2016

DNA methylation biomarkers for esophageal adenocarcinoma and precursor disease

Melissa L. Thomas
The University of Notre Dame Australia

DNA METHYLATION BIOMARKERS FOR ESOPHAGEAL ADENOCARCINOMA AND PRECURSOR DISEASE

MELISSA LEE THOMAS
BSc (with Distinction)

This thesis is presented in fulfillment of the requirements for the degree of

Doctor of Philosophy
School of Medicine
University of Notre Dame Australia

2016

Supervisor: Prof. Reginald V Lord
Co-supervisor: Prof. Susan J Clark

Statement of candidate contribution

'I hereby declare that the entirety of the work contained herein is my own and to the best of my knowledge contains no materials previously published or written by another person, except where due acknowledgement is made in the thesis. Any contribution made to the research by others is explicitly acknowledged in the thesis. I have not previously, in its entirety or in part, submitted it for obtaining any qualification in any university or other institution.'

Signed .....

Date .....2nd November 2016.....

ABSTRACT

Esophageal adenocarcinoma has one of the poorest outcomes of all solid tumors, attributable, at least in part, to lack of an early stage diagnostic test. Aberrant methylation is an early and frequent event in carcinogenesis providing an opportunity for early cancer detection. The overall aim of this thesis is to identify and validate regions of aberrant methylation as a biomarker for early detection of esophageal adenocarcinoma and the dysplastic stages of its precursor disease, Barrett's esophagus. By using well-classified patient data and stringent, quality controlled biospecimen selection for training and validation cohorts, I found regions of disease-associated aberrant methylation that are novel for esophageal carcinogenesis. With comprehensive technical and independent validation by targeted amplicon sequencing and whole genome methylation profiling of a large external validation cohort, I demonstrated potential utility of these target regions for identification of intervention requiring disease. For subsequent blood investigation, all target regions are unmethylated in peripheral blood from healthy patients and amplification assays for targeted sequencing are suitable for degraded, shorter fragments of cell-free circulating DNA in blood. I proposed a panel of three methylation biomarkers (TUBA3FP, VANGL2, ARL10) for identification of intervention requiring disease, reporting 100% sensitivity and 84.6% specificity and demonstrated biomarker application for prediction of disease progression as well as utility for monitoring disease status with treatment. I was also able to show utility for predicting the necessity of treatment for low-grade dysplasia, which is controversial in guidelines worldwide. By performing genome-wide methylation and expression profiling, as well as cancer-associated mutation screening on single tissue biopsies from all stages of the metaplasia-dysplasia-adenocarcinoma sequence, I was able to gain a more complete understanding of the genetic and epigenetic changes occurring in esophageal adenocarcinogenesis. The research presented in this thesis demonstrates that I have been able to propose potentially clinically valuable methylation biomarkers for the detection of intervention-requiring disease,

with potential application for non-invasive, high-risk population screening for identification of esophageal adenocarcinoma at an early, treatable stage.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Professor Reginald V Lord and co-supervisor, Professor Susan J Clark for their continuous support of my PhD study and related research. In particular, the support and mentorship from Professor Clark, both academically and personally, especially towards the end of my candidature was invaluable. My sincere thanks also goes to Professor Lord's gastroesophageal cancer research team and Professor Clark's epigenetics research team for all their help and support. In particular I would like to thank Dr Elena Zotenko; the project would not have been possible without her bioinformatic analyses and her continual push for me to think about how my research fits into the bigger picture. Elena's shared enthusiasm and excitement for my project was incredibly important to me.

Each section of this thesis depended on the valuable assistance of the following individuals:

- Patient Cohort and Sample Collection: Angelique Levert-Mignon, for her invaluable assistance with patient recruitment, sample collection and database upkeep. Associate Professor Yuri Bobryshev, Dr Melanie Edwards, Associate Professor Michael Buckland and Dr Min Ru Qu for their hours spent at the microscope to enable accurate pathology-based sample classification fro the study.
- Genome-wide Methylation and Expression Profiling: Dr Elena Zotenko, for her methylation and expression microarray data analyses and Dr Mark Cowley for his input regarding analysis of the new transcriptome HTA2.0 arrays.
- Target Region Validation and Characterization: Professor Sue Clark's team of scientists, in particular Dr Shamila Nair for her time spent answering my many questions about experimental design, Wenjia Qu for her help with method optimization, Ksenia Skvortsova for helping with MiSeq desktop sequencer set up (especially when the instrument was being stubborn!), and Dr Phuc Loi Luu for his pre-processing and

analysis of the resultant sequencing data. Thanks to Dr Andrew Barbour and his team at the University of Queensland / QIMR for their collaboration and kind sharing of HM450 data prior to publication, as an external validation data set for this project.

- Pan-Cancer Mutation Profiling: Dr Mark McCabe for his help in library preparation and processing of raw data for the pan-cancer mutation screening panel.

Last, but not least, I would like to thank my husband, Richard Pattison, for his support, love and valuable advice; and my friends and family who encourage and support me in all aspects of my life.

FINANCIAL SUPPORT

This study was partially funded by a National Health and Medical Research Council (NHMRC) Centre of Research Excellence (CRE) grant (Application ID APP1040947, PROBE-NET: The Progression of Barrett's Esophagus to cancer Network).

Supporting grants

During my PhD I was grateful to receive further financial support towards project reagents and consumables through the following awards:

- St Vincent's Clinic Foundation, Thelma Greig Cancer Grant. DNA methylation biomarkers for Barrett's esophagus and esophageal adenocarcinoma. CIA Reginald V Lord, Melissa Thomas (2014)
- University of Notre Dame Australia, School of Medicine Research Support Grants. Discovery and early phase validation of molecular biomarkers for gastrointestinal cancer and precursor disease. Melissa Thomas, Reginald V Lord (2014)
- St Vincent's Clinic Foundation, K&A Collins Cancer Grant. DNA methylation biomarkers: towards a diagnostic blood test for Barrett's esophagus and esophageal adenocarcinoma. Melissa Thomas, Reginald V Lord (2015)

Scholarships

- 2013 – 2016: University of Notre Dame Australia, School of Medicine PhD scholarship

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

LIST OF ABBREVIATIONS

Note: reagent and chemical abbreviations are expanded in Table 2-1.

Abbreviation	Expanded term
ARL10	ADP(adenosine diphosphate)-ribosylation factor-like 10
BE	Barrett's esophagus
BRAF	v-raf murine sarcoma viral oncogene homolog B1
BSP	Bisulphite-specific PCR
CA4	Carbonic anhydrase IV
CACNA2D2	Calcium channel, voltage-dependent, alpha 2/delta subunit 2
CADD	Combined Annotation Dependent Depletion
cfc-DNA	Cell free circulating DNA
CIMP	CpG island methylator phenotype
CNV	Copy number variation
CRE	Centre of Research Excellence
ct-DNA	Circulating tumor DNA
DE	Differentially expressed
DMR	Differentially methylated region
dsDNA	Double-stranded DNA
EAC	Esophageal adenocarcinoma
EDTA	Ethylenediaminetetraacetic acid ($C_{10}H_{16}N_2O_8$)
ELOVL5	ELOVL fatty acid elongase, family member 5
EMR	Endoscopic mucosal resection
EMT	Epithelial-to-mesenchymal transition
ES cells	Embryonic stem cells
FDR	False discovery rate
GEO	Gene expression omnibus
GEJ	Gastroesophageal Junction
GI	Gastrointestinal, for example GI cancers
GSEA	Gene set enrichment analysis
H&E	Hematoxylin and eosin histology stain
HALO	Radiofrequency ablation treatment
HGD	High grade dysplasia
HM450	Illumina's Infinium HumanMethylation450 BeadChip
HREC	Human Research Ethics Committee
HTA2.0	Affymetrix GeneChip Human Transcriptome Array 2.0
IDH1/IDH2	Isocitrate dehydrogenase-1 / -2
IGV	Integrative genomics viewer
IM	Intestinal metaplasia (transformation of epithelium into intestinal-like epithelium)
Indel	Insertion or deletion of DNA bases
ISM2	Isthmin 2
KLF7	Kruppel-like factor 7 (ubiquitous)
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LGD	Low grade dysplasia
LRRC43	Leucine rich repeat containing 43
MGMT	O-6-methylguanine-DNA methyltransferase
NHMRC	National Health and Medical Research Council
NTC	No template control
PCA	Principal component analysis
PROBE-NET	Progression of Barrett's esophagus to cancer Network
QIMR	Queensland Institute of Medical Research
R132	Arginine-132
REC	Residual embryonic cell
RefSeq	NCBI Reference Sequence
RIN	RNA integrity number
SCC	Squamous cell carcinoma

Abbreviation	Expanded term
SCOC	Short coiled-coil protein
SNP	Single nucleotide polymorphism
SPRI	Solid phase reversible immobilization
SSA	Site Specific Approval
TBE	Tris-Borate-EDTA buffer
TCGA	The Cancer Genome Atlas
TEPP	Testis, prostate and placenta expressed
TET2	Tet Methylcytosine Dioxygenase 2
TGF	Transforming growth factor
TKCC	The Kinghorn Cancer Centre, a joint facility of St Vincent's Hospital and Garvan Institute of Medical Research
Tm	Melt temperature
TNFAIP8L3	Tumor necrosis factor, alpha-induced protein 8-like 3 (also known as TIPE3)
TP63	Tumor protein 63, also known as p63
TRANK1	Tetracopeptide repeat and ankyrin repeat containing 1
TSS	Transcription start site
TUBA3FP	Tubulin alpha 3f pseudogene
VANGL2	Vang-like 2, VANGL planar cell polarity protein 2
VEP	Variant Effect Predictor
WNT	Wingless-type MMTV integration site family, encoding secreted signaling proteins, implicated in development and carcinogenesis
ZNF	Zinc finger protein

Note on spelling

Note that USA spelling has been used throughout, as most esophageal publications use USA rather than British spelling.

MANUSCRIPTS ARISING DURING MY THESIS CANDIDATURE

Maag JLV, Fisher OM, Levert-Mignon A, Kaczorowski DC, **Thomas ML**, Hussey DJ, Watson DI, Wettstein A, Bobryshev YV, Edwards M, Dinger ME, Lord RV. Novel aberrations in Barrett's esophagus and esophageal adenocarcinoma identified through whole transcriptome sequencing. Manuscript submitted to Nature Genetics, awaiting review.

For this publication I co-ordinated the study, created the study design and selected specimen cohorts for sequencing.

Levert-Mignon A, Bourke MJ, Lord SJ, Taylor AC, Wettstein AR, Edwards M, Botelho NK, Sonson R, Jayasekera C, Fisher OF, **Thomas ML**, Macrae F, Hussey DJ, Watson D, Lord RV. Changes in gene expression of neo-squamous mucosa after endoscopic treatment for dysplastic Barrett's esophagus and intramucosal adenocarcinoma. United European Gastroenterol. J. 2016, May 19. Doi:10.1177/2050640616650794.

For this publication I provided input regarding analyses and manuscript preparation.

Fisher OM, Levert-Mignon AJ, Lord SJ, Botelho NK, Freeman AK, **Thomas ML**, Falkenback D, Wettstein A, Whiteman DC, Bobryshev YV, Lord RV. High Expression of Cathepsin E in Tissues but not Blood of Patients with Barrett's Esophagus and Adenocarcinoma. Ann Surg Oncol. 2015, July; 22(7):2431-8. doi: 10.1245/s10434-014-4155-y.

For this publication, I was responsible for methods troubleshooting as well as manuscript preparation.

Whiteman DC, Appleyard M, Bahin FF, Bobryshev YV, Bourke MJ, Brown I, Chung A, Clouston A, Dickins E, Emery J, Eslick GD, Gordon LG, Grimpen F, Hebbard G, Holliday L, Hourigan L, Kendall BJ, Lee EY, Levert A, Lord RV, Lord SJ, Maule D, Moss A, Norton I, Olver I, Pavey D, Raftopoulos S,

Rajendra S, Schoeman M, Singh R, Sitas F, Smithers BM, Taylor A, **Thomas ML**, Thomson I, To H, von Dincklage J, Vuletich C, Watson DI and Yusoff IF. Australian clinical practice guidelines for the diagnosis and management of Barrett's Esophagus and Early Esophageal Adenocarcinoma. J. Gastroenterol. Hepatol. 2015, May; 30(5):804-20. doi: 10.1111/jgh.12913.

Authors in alphabetical order. I was responsible for review and evaluation of literature to answer the question 'Are there biomarkers for the diagnosis (presence) of Barrett's esophagus?'. This work was primarily done for Cancer Council Australia's online Clinical Practice Guidelines for the Diagnosis and Management of Barrett's esophagus and early Esophageal Adenocarcinoma. Overall findings were summarized in this publication.

Fisher OM, Levert-Mignon AJ, Lord SJ, Lee-Ng KKM, Botelho NK, Falkenback D, **Thomas ML**, Bobryshev YV, Whiteman DC, Brown DA, Breit SN, Lord RV. MIC-1/GDF15 in Barrett's Esophagus and Esophageal Adenocarcinoma. Br. J. Cancer. 2015, Apr 14; 112(8)1384-91. doi: 10.1038/bjc.2015.100.

I was part of the team responsible for sample and data collection, as well as specimen selection for the study, sample extraction and processing.

CHAPTER 1: INTRODUCTION

1.1 Introduction

Cancer is no longer thought of as purely a genetic disease, with epigenetic alterations proving ubiquitous across nearly all human cancers[1]. In some cases, such as retinoblastoma and rhabdoid tumor, epigenetic change has even been identified as the principal driver of carcinogenesis[2, 3]. Altered DNA methylation occurs early and frequently in cancer development with some changes already present in precursor lesions[4], making it an attractive choice for a diagnostic biomarker.

Esophageal adenocarcinoma (EAC) is the sixth leading cause of cancer death in the western world[5]. Furthermore, EAC incidence rates have risen faster than that of any other cancer in western countries over the past several decades; thus research into improved early detection, treatment and prevention is of high priority. Effective intervention for EAC is possible and often curative with diagnosis at an early stage[6, 7], but most patients are diagnosed at a late, incurable stage, contributing to the poor prognosis and high case fatality ratio. One of the major challenges associated with EAC is early disease identification. DNA methylation biomarkers have been shown to have clinical application for early detection in cancers such as colorectal and lung[1], and have similar potential for EAC.

This introductory chapter aims to provide a summary of research in the field of epigenetic aberrations in EAC development, focusing on how these changes could be used as biomarkers for early disease detection.

1.2 Epigenetics and cancer

Epigenetics refers to heritable changes that do not alter the underlying DNA sequence, including key processes such as DNA methylation, chromatin modifications and nucleosome positioning. Mammalian DNA methylation involves the addition of a methyl group (CH_3) to the 5^{th} carbon position on a

cytosine base of a CpG dinucleotide (cytosine followed directly by guanine) in the DNA sequence. Eighty to ninety percent of CpG sites in human DNA are methylated, however promoter regions of mammalian genes are often GC-rich (known as CpG islands) and entirely unmethylated, correlating with increased transcriptional activity[8]. The presence of methylation in promoter regions is transcriptionally repressive, whereas a lack of methylation allows for active transcription of the associated gene[9]. DNA methylation is widely regarded as the most stable and informative epigenetic mark for understanding gene expression and cell differentiation[10].

Chromatin is a complex of DNA and protein (histone) that dictates DNA structure in eukaryotic cell nuclei. Histones act as a spool around which DNA can wind and when modified, can influence chromatin arrangement and ultimately, gene transcription (Figure 1-1). Condensed chromatin (heterochromatin) is typically not transcribed, whereas the less compact form of chromatin (euchromatin) is active and the associated DNA can be transcribed.

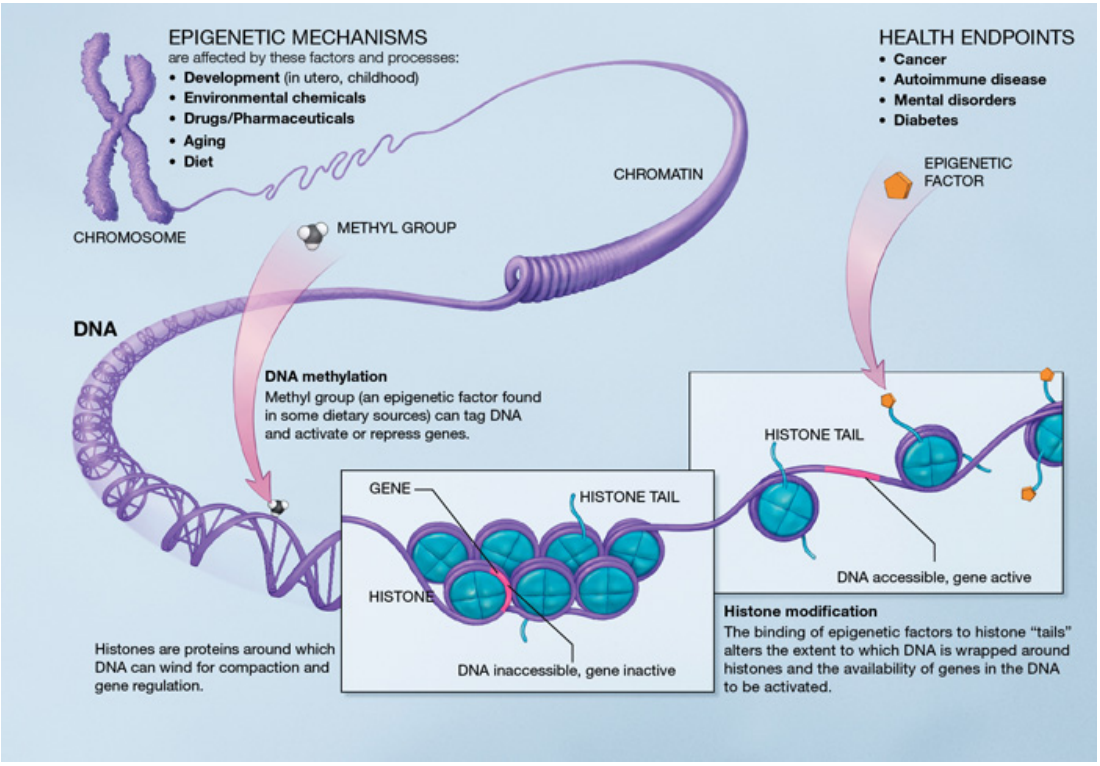


Figure 1-1: Gene expression is influenced by epigenetic mechanisms such as DNA methylation and histone modification. Figure courtesy of the National Institutes of Health (http://commonfund.nih.gov/epigenomics/figure.aspx)

Aberrant DNA methylation plays an important role in carcinogenesis by either inactivation of usually active tumour suppressor genes (disease-associated hypermethylation), or activation of normally inactive oncogenes (demethylation, or disease-associated hypomethylation). Gene silencing associated with promoter CpG island DNA hypermethylation has been identified not only in many cancers, but also precursor lesions (Figure 1-2). Early stage aberrant DNA methylation may indicate potential disease progression and subsequent carcinogenesis and thus may be useful as a biomarker for early cancer detection[11].

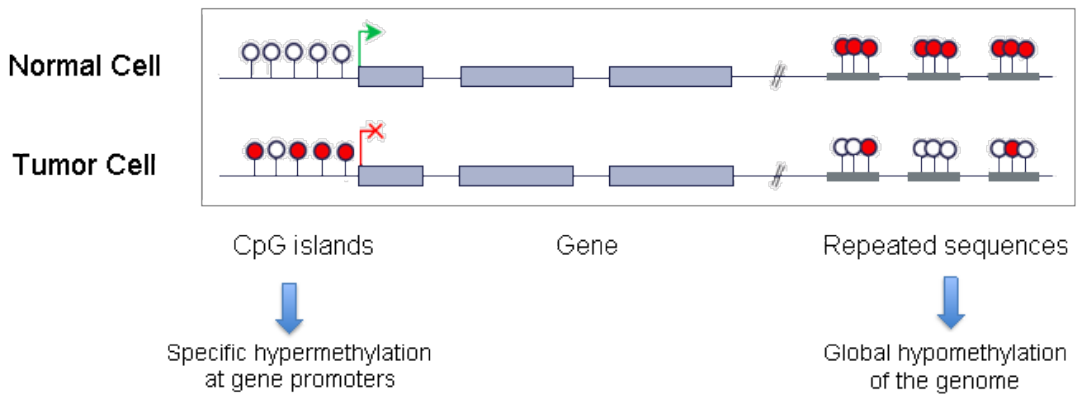


Figure 1-2: Aberrant DNA methylation in cancer. Tumor cells show increased methylation in promoter regions of specific loci, but a global decrease in DNA methylation at non genic regions, for example in repeated sequences. White lollipop: unmethylated CpG site, red lollipop: methylated CpG site.

The field of epigenetics is one of the most promising and rapidly expanding areas in biomarker research, with mechanisms such as DNA methylation and histone modification likely to be some of the earliest events in carcinogenesis[12]. Importantly, epigenetic disruption has been found to precede genetic change, further supporting the use of epigenetic mechanisms such as DNA methylation as early detection biomarkers. Recent advances in technology have lead to a deeper understanding of

epigenetic processes occurring during carcinogenesis and a resultant focus on epigenetic alterations (in particular, aberrant DNA methylation) as biomarkers for early cancer detection, prognosis and disease monitoring [8, 12].

1.2.1 Methylation biomarkers for early cancer detection

One of the biggest challenges in oncology is early disease detection. Major efforts are underway to identify biomarkers that enable detection of early-stage, treatable, and potentially curable carcinogenic change rather than advanced, often incurable cancer. Epigenetic alterations are now well recognized as a key step in carcinogenesis, with promoter hypermethylation of tumour suppressor genes showing promise as early-stage biomarkers in a number of different cancer types[13, 14].

Aberrant methylation is specific, stable and detectable in peripheral blood, making it an attractive cancer biomarker. Despite its suitability as a cancer detection aid, being detectable in blood, stool, urine and solid biopsy[15-23], and despite the plethora of potential DNA methylation biomarkers proposed for a wide variety of cancer types, only minimal DNA methylation biomarkers have been clinically implemented. Commercial tests are currently available for just a handful of DNA methylation biomarkers for both diagnostic purpose and cancer therapy treatment decisions[1].

MDxHealth's PredictMDx™ for Glioblastoma tests for methylation of the DNA repair gene, O-6-methylguanine-DNA methyltransferase (MGMT) in glioblastoma (the most common and aggressive malignant primary brain tumour in humans) tissue, and uses results to identify patients most likely to respond to alkylating agent treatment. MGMT methylation was also shown to have prognostic value in glioblastoma, predicting significant increase in both overall and progression-free survival[24]. Epigenomics' Epi proLung® BL assay detects methylated Short Stature Homeobox 2 (SHOX2) in bronchial washings from patients with suspected lung carcinoma, introduced into clinical practice as a routine diagnostic aid in a German hospital in 2012. Very recently, SHOX2 methylation has been investigated in blood, with

promising results for not only diagnostic purposes but also therapy monitoring[25, 26].

The most advances in commercially available methylation biomarkers are in the field of colorectal cancer. Exact Sciences Cologard™ measures methylated BMP3 and NDRG4 (as well as point mutations in KRAS, haemoglobin and β-actin (as a reference)) in stool as a non-invasive screening test, intended as an adjunct to colonoscopy, reporting 92% sensitivity for detection of colorectal cancer, 42% for advanced precancerous lesion and 87% specificity[27]. Further studies have extended these findings, reporting 85-100% detection of colorectal cancer and 53-83% for high-grade dysplasia and adenoma[28-31]. Commercial blood tests for the early detection of colorectal cancer have been launched by Epigenomics (Epi proColon®), Quest Diagnostics (ColoVantage®) and Abbott Molecular (RealTime mS9); all measuring the methylation status of Septin 9 (SEPT9). In Australia, Quest Diagnostics have teamed with Clinical Genomics who have recently launched regional pilot ColoVantage Plasma testing, which examines methylated BCAT1 and IKZF1 in patient plasma for early colorectal cancer diagnosis. Methylation in these two targets has also shown utility for disease recurrence, with supporting data presented at the American Society of Clinical Oncology Gastrointestinal Cancers (ASCO GI) Symposium in San Francisco on 23 January 2016.

Over the past two decades, the use of DNA methylation biomarkers for cancer diagnosis has been assessed, with many alterations identified as potential biomarkers but only few translated to clinical use[32]. One of the major obstacles limiting clinical implementation appears to be variability in reported accuracy in subsequent studies. The use of comprehensive validation, large independent cohorts, and standardization of methylation assessment may help reduce this variability. Another key issue for clinical uptake is standardization of quantitative methodology, ensuring multi-laboratory reproducibility[1].

1.3 Barrett's esophagus and esophageal adenocarcinoma

Barrett's esophagus (BE) is the replacement of normal squamous epithelium with columnar, goblet-cell containing epithelium in the distal esophagus, in response to long-standing gastro-esophageal reflux disease (GERD)[33]. It is a relatively common disorder, with an estimated incidence of 1-2% of the adult population[34], but often or even usually, remains undiagnosed. Barrett's esophagus arises in the distal esophagus, extending proximally from the gastroesophageal junction (GEJ) and is measured using Prague Criteria (Figure 1-3).

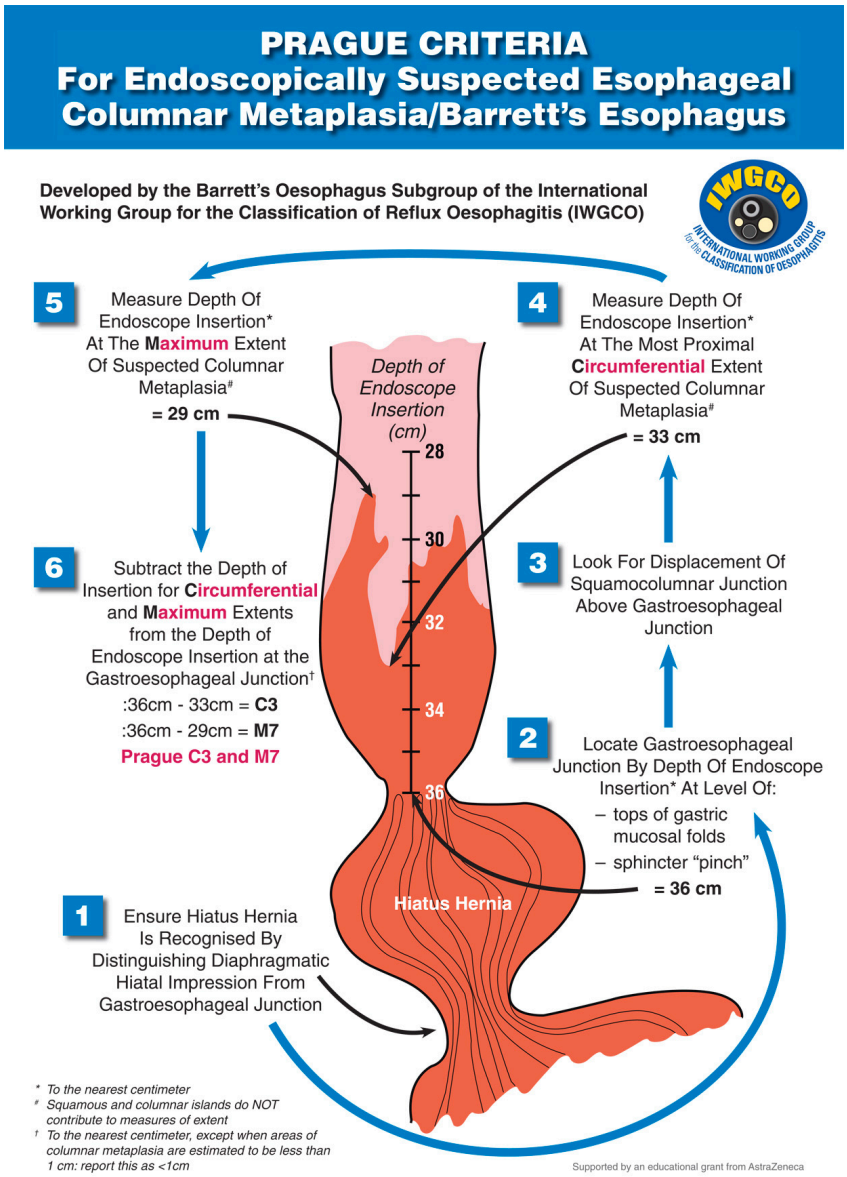


Figure 1-3: Barrett's esophagus extent is quantified using Prague Criteria. Prague Criteria specifies a circumferential extent and maximal extent, measured in centimeters from the defined GEJ. For example, BE (C3 M7) is Barrett's mucosa extending circumferentially for 3cm, with maximal tongues as far as 7cm from the GEJ. Figure courtesy of the International Working Group for the Classification of Reflux Oesophagitis (IWGCO), http://www.iwgco.net. The criteria and their validation is outlined by Sharma et al, 2006[35].

The chief significance of BE lies in its role as precursor disease to esophageal adenocarcinoma (EAC)[36], which arises following a multistage metaplasia-dysplasia-adenocarcinoma progression (as outlined in

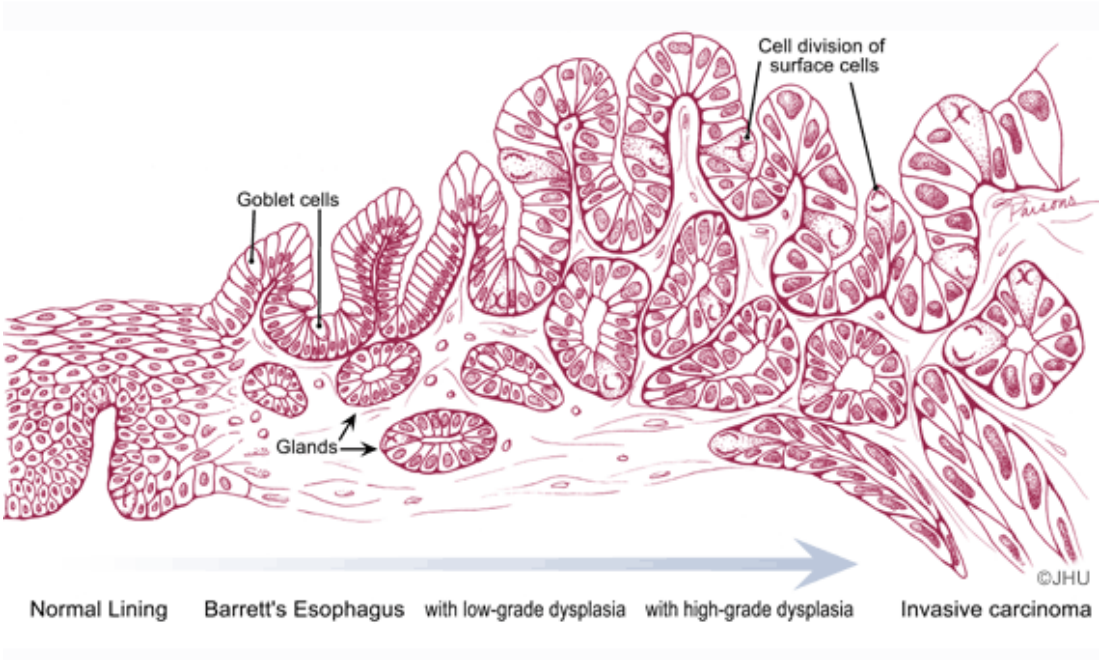


Figure 1-4). The presence of BE has been reported to increase risk of EAC 30- to 125-fold[37, 38]. However, only a small proportion of BE sufferers (estimated as less than 0.5% per year[39]) will progress from non-dysplastic disease, highlighting the need for stratification of BE patients to identify those at risk of dysplastic progression.

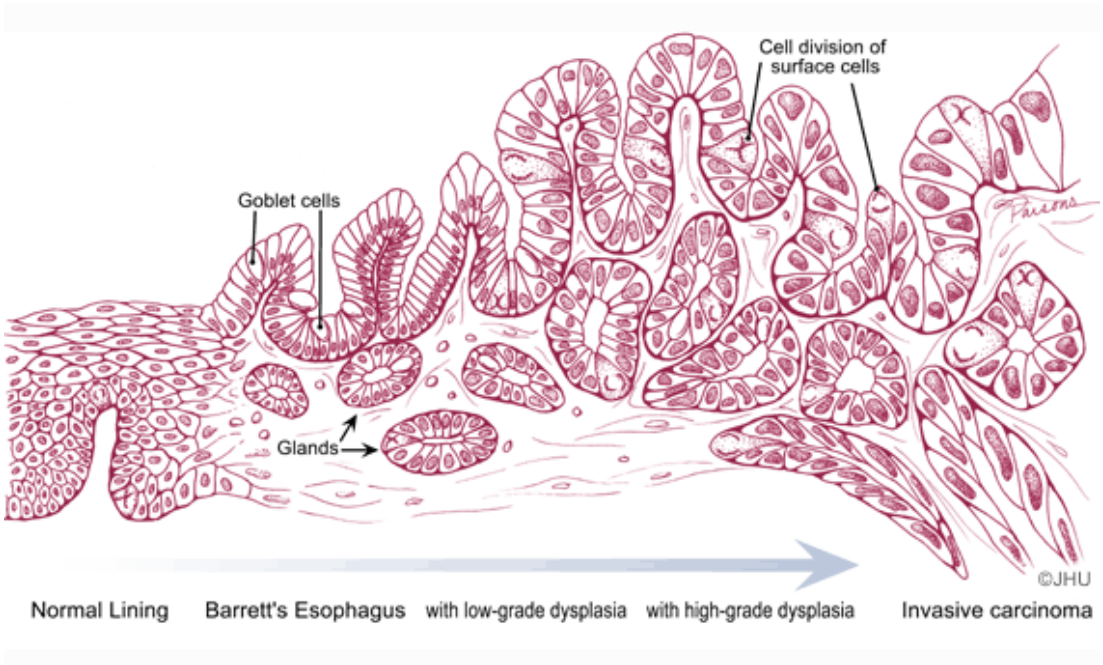


Figure 1-4: The multistage development of esophageal adenocarcinoma: from normal healthy squamous epithelium through metaplasia, to dysplasia and finally adenocarcinoma. Image courtesy of John Hopkins Pathology (http://pathology2.jhu.edu/beweb/cancer.cfm).

Of all solid tumors, EAC has one of the poorest outcomes, with 5-year survival rates as low as 14%[40, 41], however a three times higher survival rate has been reported when associated with early diagnosis as opposed to metastatic disease at diagnosis[42].

1.3.1 The need for an alternative to histology-based diagnosis

Current clinical methods for BE diagnosis require endoscopic identification of columnar esophageal mucosa and histological confirmation of the presence of intestinal metaplasia with mucin-containing goblet cells. This combined endoscopic and histology-based method, also used for EAC diagnosis, presents a number of challenges in the detection of early stage disease. First and foremost, these methods are invasive and not viable for even high-risk population screening (such as Caucasian males aged over 40 with longstanding reflux) due to significant expense to the health care system and limited availability of specialist time and facilities. Thus with endoscopy-histology based diagnosis, disease is often detected at a late-stage,

contributing to the extremely poor prognosis associated with esophageal adenocarcinoma.

Secondly, large numbers of biopsies must be taken to reduce sampling error and ensure that the 'worst case' diagnosis isn't missed. Even with these precautions, low-grade dysplasia (LGD) is frequently not detected or missed in diagnosis. Poor diagnostic reproducibility is common for patients diagnosed with LGD, with subsequent endoscopy often reporting the presence of non-dysplastic Barrett's mucosa only, despite no treatment being received for LGD.

Current histology-guided clinical care for BE/EAC is problematic, not only due to endoscopic sampling error but also histological misclassification, due to poor sample preparation, biological heterogeneity or other artefacts[43]. Furthermore, pathology-based dysplasia assessment has been shown to be highly variable, with substantial inter- and intra-observer variation reported. Late stage detection and poor prognosis of EAC (5-year survival rates are typically <15%[44]) are potentially partly attributable to ineffective endoscopic/histologic diagnosis methods and may be overcome by identification of accurate, robust biomarkers occurring early in the development of EAC.

1.3.2 Epigenetic regulation of genes in EAC carcinogenesis

It is well established that promoter hypermethylation is a frequent mechanism of transcriptional inactivation and gene silencing of tumor suppressor genes[45, 46] and there is significant evidence that these epigenetic alterations play an important role in EAC carcinogenesis[38, 47]. Genome-wide analysis reveals that global hypomethylation is the dominant change in BE development, followed by a much smaller wave of selective promoter hypermethylation in the progression to adenocarcinoma[48, 49].

In a comprehensive review of epigenetic alterations in BE pathogenesis to EAC by Argawal et al in 2012, 35 genes were reported to be aberrantly methylated in BE and EAC, all originating from single-locus studies[48]. Genes

observed to be aberrantly methylated in EAC tissues include AKAP12, APC, CDH13, DAPK1, GPX3, MGMT, p16/CDKN2A, RUNX3, SFRP1, TERT, TIMP3 and WIF-1[48]. To date the only studies which have addressed identification of dysplastic BE progression are not convincing in their specificity. For example, APC, TIMP3 and TERT promoters were hypermethylated in 90-100% of Barrett's mucosa from patients who had progressed to EAC but specificity was poor, ranging from 15 to 35%[50].

In the only genome-wide EAC methylation study to date, Krause et al (2016)[49] reported on the lack of separation between BE and EAC cohorts in unsupervised clustering, indicating that aberrant methylation is an early event in metaplasia-dysplasia-neoplasia progression (Figure 1-5). Only mentioned in passing and not the focus of this study, just 2024 CpG sites were identified as differentially methylated between EAC and BE (in comparison to the 52,590 CpG sites differentially methylated in EAC compared to normal epithelium). This smaller subset is the focus of my thesis; the differentially methylated regions with clinical utility for early detection EAC screening.

Figure 1-5: Aberrant methylation occurs early in EAC carcinogenesis as evidenced by differentially methylated probe and gene counts in EAC and BE (compared to normal esophageal mucosa) identified by Krause et al[49] in the first genome-wide EAC methylation study to date. FDR < 0.01 and beta value difference ≥ 0.20.

1.3.3 Proposed diagnostic biomarkers

Lack of early diagnostic screening for Barrett's esophagus and esophageal adenocarcinoma is not due to lack of research, with numerous studies

investigating alternatives to current combined endoscopic/histologic methods. Despite the plethora of studies performed, none have replaced or been universally implemented as a supplement to current clinical practice.

1.3.3.1 Barrett's esophagus diagnostic biomarkers

Since 2013, I have been a member of Cancer Council Australia's Barrett's Esophagus and Early Esophageal Adenocarcinoma Working Party, responsible for evaluating current literature and collating evidence into an online Wiki: Clinical practice guidelines for the diagnosis and management of Barrett's esophagus and early esophageal adenocarcinoma[51]. More specifically, I was responsible for review and evaluation of literature (23 articles) to answer the question 'Are there biomarkers for the diagnosis (presence) of Barrett's esophagus?' (http://wiki.cancer.org.au/australia/Clinical_question:Are_there_biomarkers_for_the_diagnosis_(presence)_of_BO%3F). Overall findings were summarized in a publication in the Journal of Gastroenterology and Hepatology in May 2015[52]; a full copy of this paper is included in the final section of this thesis: Reprints of publications arising during my thesis candidature.

In brief, I found that numerous biomarkers had been proposed to aid in the diagnosis of Barrett's esophagus, including tissue and serum markers as well as evaluation of a non-endoscopic capsule sponge device paired with immunohistochemistry for Trefoil factor 3 (TFF3). None of the studies provided sufficient evidence for recommendation as a supplement to replace current clinical practice due to high risk of bias in study design, wide variation in accuracy estimates across studies and no evidence of comparison with current standard practice.

1.3.3.2 Esophageal adenocarcinoma diagnostic biomarkers

In 2009, a panel of eight DNA-methylation biomarkers for predicting the risk of progression from BE to esophageal adenocarcinoma (p16, RUNX3, HPP1, NELL1, TAC1, SST, AKAP12 and CDH13) was trialed in a multi-center, retrospective, double-blinded validation study[38], reporting 50% sensitivity

(predicting half the HGD and EAC incidence) when specificity was set to 0.9 (10% of non-progressors falsely predicted as progressors). Three of these biomarkers originated from a real-time quantitative methylation specific PCR (MSP) study in 2005[53], based on previous studies from as early as 1997-2002 [47, 54-58]). Interestingly, APC and TIMP3 were also tested by Schulmann in this study and not selected as possible predictive methylation markers for increased risk of EAC; but later proposed (alongside TERT) as a possible panel for this purpose by Clement et al in 2006[50]. The further five biomarkers comprising the eight gene panel were selected based on MSP interrogation by Jin et al. in 2007-2008[59-63]. Interestingly, all eight biomarkers were originally selected based on their status as common methylation-inactivated tumor suppressor genes in other forms of human neoplasia, rather than a targeted screening approach specific for EAC and its precursor disease BE.

This panel of eight DNA methylation biomarkers to predict the risk of progression proposed by Stephen Meltzer and his team at John Hopkins University[38] has been launched commercially in 2013. A significant limitation of the Meltzer laboratory test is its reliance on endoscopic tissue biopsies.

In a 2013 study by Alvi et al[64], a new four gene DNA methylation panel (SLC22A18, PIGR, GJA and RIN2) was put forward as an adjunct to histopathology to enable stratification of patients into three risk groups based on the number of genes methylated. However this work is still in research phase and requires further validation before it can be considered for clinical application. HumanMethylation27 BeadChip® (HM27) microarrays were used for discovery (22 BE, 2 duodenum and 24 EAC samples), with selected genes validated using pyrosequencing[40, 64]. In the field of esophageal cancer, more comprehensive Infinium HumanMethylation 450K (HM450) arrays have been used to investigate potential biomarkers for the early diagnosis of esophageal squamous cell carcinoma (SCC)[65], however the same cannot be said for EAC and its precursor disease BE. Searching GEO datasets platform GPL13534 (publicly available HM450 data) AND esophagus returns only GSE72874 (released 1st April 2016, by Krause et al[49]; these data were used as an external validation data set for this thesis). However this single HM450

data set for EAC does not include any dysplastic (low or high-grade) samples. Thus, my study is novel in its analysis of the full metaplasia-dysplasia-adenocarcinoma progression sequence (including both low and high grade dysplastic mucosa) using HM450 technology.

Significant epigenetic changes occur during EAC carcinogenesis, however much of the focus of biomarker research to date has been directed towards aberrant methylation observed in EAC when compared to normal mucosa. For many cancers this 'tumor v normal' approach will work, but for EAC the situation is more complicated. Much of the aberrant hypermethylation in EAC is already present in its precursor disease, Barrett's esophagus[49]. However, just 0.5% of BE sufferers will ever progress, and BE itself is otherwise asymptomatic and benign. Thus, for a clinically viable early-EAC biomarker, it is important to focus on aberrant methylation present in dysplastic BE and adenocarcinoma that is NOT present in benign BE. This may be a significant contributing factor to the lack of transfer from research to clinical use that EAC biomarkers have experienced to date. Further confounding the situation, there has always been significant difficulty surrounding the diagnosis and treatment of low-grade dysplasia and thus it has generally been excluded from research studies. This is unfortunate as identification at this early treatable stage is key to reducing current poor prognosis of EAC. My thesis aims to address this by inclusion of low- (and high-) grade dysplastic mucosa and a focus on hypermethylation biomarkers present in dysplasia and EAC, using methylation in both BE and normal healthy mucosa as baseline.

1.3.4 The future: blood-based biomarkers

Blood is one of the most accessible biological fluids; it comes into close contact with all tissues and can carry disease specific biomarkers. From a practical perspective, blood analysis is already a well-established clinical routine; the infrastructure is already in place, therefore minimal effort is required to integrate validated blood biomarkers into clinical practice. Furthermore, blood-based biomarkers have the capacity to be easily reassessed over time, essential for monitoring disease progression and to

detect emerging recurrence. Blood tests are ideal to meet the current deficiencies surrounding EAC diagnosis: reproducible, easily integrated into healthcare management systems and a cost-effective method for screening high-risk populations[66]. Development of blood-based biomarkers has been shown to potentially improve detection rates in cancers such as ovarian, prostate, colorectal and gastric[67-70] and would stand to do the same for EAC.

Cell-free circulating DNA (cfc-DNA), present in human blood plasma is known to contain DNA methylation patterns characteristic of disease occurring elsewhere in the body[11]. The origins of cfc-DNA remain unclear; both passive release during cell death and active secretion from proliferating cells have been postulated. Importantly, disease diagnosis and monitoring has been shown to be possible using cfc-DNA from blood plasma as it contains mutations and DNA methylation patterns characteristic of disease[11, 71, 72].

The challenge in using cfc-DNA from blood plasma for biomarker discovery is its low concentration (present in ng/mL range); only a few genes can be interrogated. One solution to this problem is to start with well-characterised tumour tissue for biomarker discovery then transition to blood once potential biomarkers have been uncovered. Traditionally, the vast majority of attempts to develop diagnostic biomarkers have been developed this way. For example, Epigenomics Inc. have developed a cfc-DNA-based biomarker, SEPT9, for colorectal cancer using this approach[23]. However, the transition from tissue-identified putative biomarkers to blood plasma is not straight-forward[11]. For example, methylated DNA analysis in ovarian adenocarcinoma in tumour tissue and in cfc-DNA from plasma resulted in methylation patterns that were similar but not identical[73]. It has also been shown that independently confirmed tumour tissue-derived biomarkers (in hepatocelluar carcinoma) did not perform equally as well in plasma[74].

With regards to EAC, there is some evidence of tissue-identified markers showing value as blood biomarkers. Tachykinin-1 (TAC1) promoter hypermethylation was examined in subjects with matched tissue and plasma

samples at various stages of neoplastic evolution by Jin et al[60]. Some level of tissue-blood correlation was observed with approximately 44% matched TAC1 methylation status observed between tissue and plasma. However, approximately 44% of patients were observed to have TAC1 hypermethylation in tissue but not in plasma and approximately12% were observed to have TAC1 hypermethylation in plasma but not in tissue[60]. Potential methylation blood biomarkers DAPK (death-associated protein kinase) and APC (adenomatous polyposis coli gene) have been examined in EAC with variable results. Methylated DAPK and methylated APC were detected in the plasma of 61% of EAC patients by Hoffmann et al in 2009[75]. However, Kawakami et al showed that hypermethylated APC was detectable in not only 92% of EAC patients but also 40% of BE patients[76], an observation with significant impact on clinical utility of the biomarker.

BE with high-grade dysplasia (HGD) and EAC are the later stages of the Barrett's metaplasia-dysplasia-adenocarcinoma sequence at which intervention by either endoscopic or operative treatment is mandated in patients who are fit for such intervention[52]. The discovery of a biomarker panel for intervention-requiring disease with potential blood test application would provide an easily implemented, and potentially cost-effective test for high-risk population screening, enabling detection of in-situ and invasive malignancy at an early, curable stage, which would ultimately lead to improved prognosis and lower incidence of incurable disease[77].

1.4 Conclusions

This chapter has presented a summary of research in the field of epigenetic aberrations in the development of esophageal adenocarcinoma and how these changes could be used as biomarkers for early disease detection, addressing the need for an alternative to current endoscopy-histology methods.

Epigenetic mechanisms, such as DNA methylation and histone modification, are some of the earliest events in carcinogenesis and thus lend themselves

to clinical application as diagnostic biomarkers for early cancer detection. This has been shown by clinically implemented methylation biomarker tests such as Epi proLung® (lung carcinoma diagnostic), PredictMDx™ (glioblastoma treatment), Cologard™ (colorectal cancer diagnostic, adjunct to colonoscopy) as well as Epi proColon®, RealTime mS9 and ColoVantage® (colorectal cancer diagnostic blood testing).

Current clinical methods for BE and EAC diagnosis require histopathologic examination of tissue biopsies taken during endoscopic examination, which is challenging on a number of levels and a strain on the health care system due to limited availability of specialist time and facilities. Significant epigenetic changes occur during EAC carcinogenesis, however no biomarkers have been implemented in clinical use to date. This may be due in part to the complicated nature of the disease (precursor BE is benign, with a progression rate of just 0.5%), and the tumor-normal focus of much research to date, rather than using a BE baseline and including challenging low-grade dysplastic samples to ensure true early detection biomarkers are identified.

There is an urgent need to introduce a non-invasive, diagnostic test for early identification of intervention-requiring disease to significantly alter clinical outcomes for this highly fatal disease. This study aims to take the first steps, by using well-classified patient data and stringent, quality controlled biospecimens to perform genome-wide methylation and expression profiling on tissue samples over the full metaplasia-dysplasia-adenocarcinoma sequence of EAC development. With comprehensive technical and independent validation by targeted amplicon sequencing and whole genome methylation profiling of a large external validation cohort, I aim to ensure robustness and the highest chance of success in moving research to clinical implementation. Importantly, for anticipated subsequent blood investigation, all target regions will be selected to be unmethylated in peripheral blood from healthy patients and amplification assays for targeted sequencing will be suitable for the degraded, shorter fragments of cell-free circulating DNA in blood.

1.5 Hypotheses, aims and study design

1.5.1 Hypotheses underpinning study design

Several hypotheses were made underlying the study design for this thesis. Firstly, that aberrant methylation in human esophageal tissue biopsies can be used to discover disease-specific methylation changes in the development of esophageal adenocarcinoma, possibly novel discoveries if samples from *all* phases of the metaplasia-dysplasia-adenocarcinoma sequence are examined. Secondly, that altered methylation status in diseased tissue is reflected and measurable in cell-free circulating DNA (cfc-DNA) in blood. With these hypotheses, this thesis aims to identify clinically valuable, potential biomarkers for the early identification of esophageal adenocarcinoma, with a view to future blood test application.

1.5.2 Aims of this thesis

The overall aim of this thesis is to identify aberrant methylation in EAC carcinogenesis with potential blood biomarker application for the early identification of intervention requiring disease.

Specific aims are:

Aim 1: To use well-classified patient data and stringent, quality-controlled biospecimen selection for training and validation cohort assembly; performing global genome-wide methylation and expression profiling at all stages of the metaplasia-dysplasia-adenocarcinoma sequence to uncover novel disease-associated aberrant methylation with potential clinical application.

Aim 2: To comprehensively validate and characterize disease-associated differentially methylated target regions indicative of early esophageal adenocarcinoma, concurrently initiating assay development with direct transferability to blood biomarker investigation.

Aim 3: To evaluate methylation biomarker performance with regard to diagnostic utility and application for disease progression prediction and disease monitoring during treatment.

Aim 4: To examine cancer-associated mutational load at all phases of disease progression, and use this information to gain a more complete understanding of genetic and epigenetic changes occurring in the development of esophageal adenocarcinoma.

From these findings, I hope to propose potentially clinically valuable methylation biomarkers for the early detection of esophageal adenocarcinoma.

1.5.3 Study Design

The study is based on well-defined patient cohorts and comprehensively classified samples and can be broken down into four phases that reflect the four aims of the thesis (Figure 1-6). It follows the well-established discovery and validation model proposed by Pepe et al[78] for biomarker development for the early detection of cancer, aiming to identify leads for potentially clinically applicable biomarkers and prioritizing these identified leads.

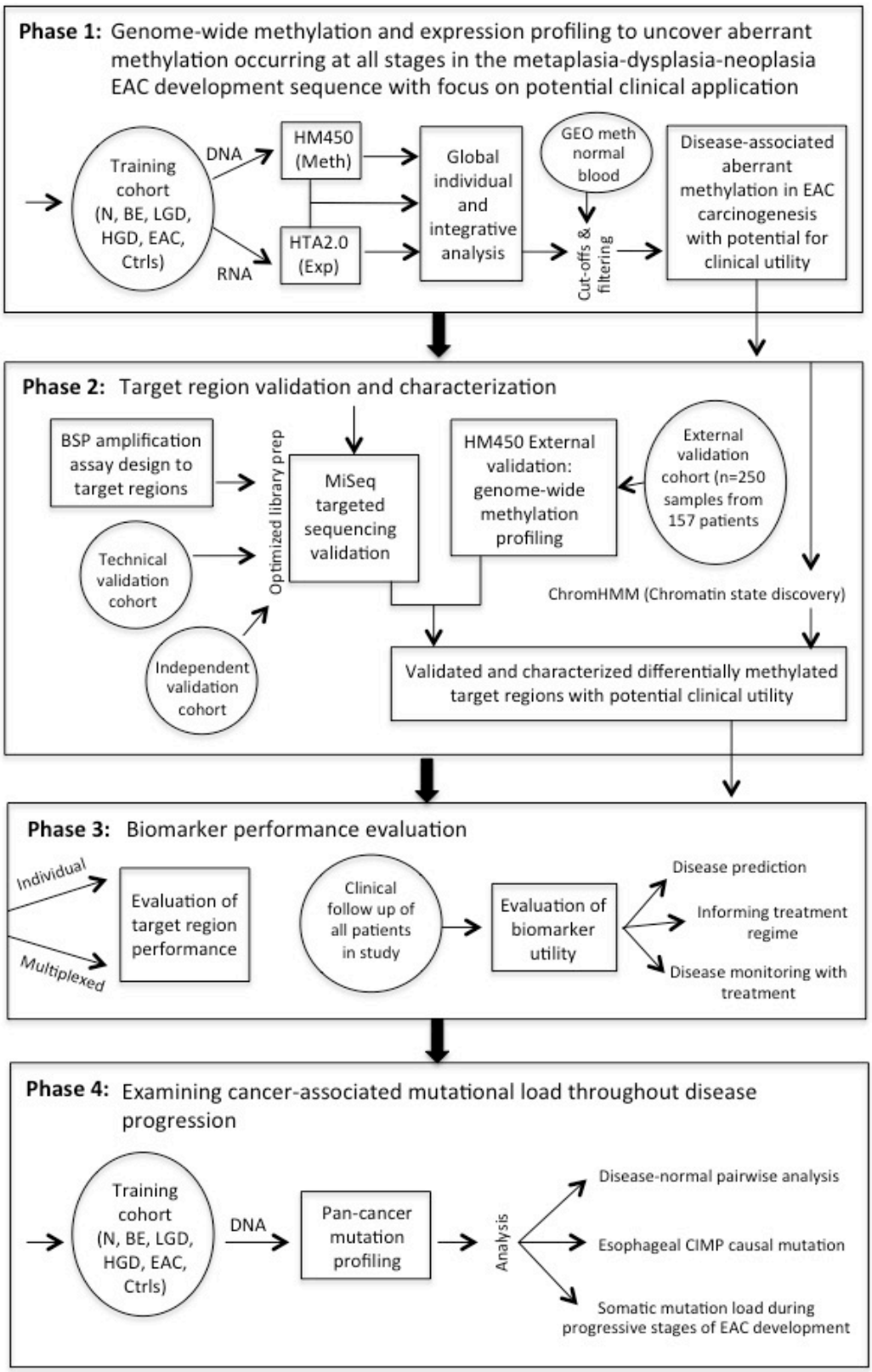


Figure 1-6: Study design schematic. This thesis aims to identify aberrant methylation in EAC carcinogenesis for the early identification of intervention requiring disease using a genome-wide epigenetic and genetic approach to

gain a complete understanding of changes occurring throughout the metaplasia-dysplasia-adenocarcinoma progression.

CHAPTER 2: MATERIALS AND GENERAL METHODS

2.1 Materials

2.1.1 Reagents and chemicals

Reagents and chemicals used for this study are summarized in Table 2-1.

Table 2-1: List of reagents and chemicals

Reagent/Chemical Name	Supplier	Supplier Location
2-Mercaptoethanol ($HSCH_2CH_2OH$)	Sigma-Aldrich	St Louis, MO, USA
2-Propanol ($(CH_3)_2CHOH$)	Sigma-Aldrich	St Louis, MO, USA
4',5'-Dibromo-2',7'-dinitroflourescein disodium salt (Eosin B), ($C_{20}H_6N_2O_9Br_2Na_2$)	Agilent Technologies	Santa Clara, CA, USA
Agarose UltraPure™ 1000	Life Technologies	Carlsbad, CA, USA
Boric acid (H_3BO_3)	Sigma-Aldrich	St Louis, MO, USA
Chloroform-isoamyl alcohol mixture 24:1, ≥99.5% for molecular biology	Sigma-Aldrich	St Louis, MO, USA
CpGenome Universally Methylated DNA	Merck Millipore	Billerica, MA, USA
Custom DNA oligonucleotides, lyophilized, 25nmole	Integrated DNA Technologies, Inc.	Coralville, IA, USA
Dimethyl sulfoxide (DMSO), ($(CH_3)_2SO$)	Sigma-Aldrich	St Louis, MO, USA
Ethanol, 200 proof (absolute) for molecular biology (CH_3CH_2OH)	Sigma-Aldrich	St Louis, MO, USA
Ethylenediaminetetraacetic acid ($(HO_2CCH_2)_2NCH_2CH_2N(CH_2CO_2H)_2$)	Sigma-Aldrich	St Louis, MO, USA
Formalin solution, neutral buffered, 10%	Sigma-Aldrich	St Louis, MO, USA
Hematoxylin ($C_{16}H_{14}O_6 \cdot xH_2O$)	Agilent Technologies	Santa Clara, CA, USA
Human Genomic DNA from blood, in 10mM Tris HCl, 1mM EDTA, pH 8.0	Roche Diagnostics	Mannheim, GER
Magnesium chloride ($MgCl_2$)	Life Technologies	Carlsbad, CA, USA
NaOH, molecular biology grade, 1.0N	Sigma-Aldrich	St Louis, MO, USA
Nuclease-free water (H_2O)	Qiagen	Hilden, GER
PhiX Control v3, 10nM	Illumina, Inc.	San Diego, CA, USA
Polysorbate 20, aka Tween 20 ($C_{58}H_{114}O_{26}$)	Sigma-Aldrich	St Louis, MO, USA
Proteinase K, 20mg/mL	Qiagen	Hilden, GER
RNA*later*® RNA stabilization reagent	Life Technologies	Carlsbad, CA, USA
Sodium Dodecyl Sulfate ($NaC_{12}H_{25}SO_4$)	Applied Biosystems	Foster City, CA, USA

Reagent/Chemical Name	Supplier	Supplier Location
SYBR® Safe DNA Gel Stain	Life Technologies	Carlsbad, CA, USA
SYTO® 9 Green Fluorescent Nucleic Acid Stain	Life Technologies	Carlsbad, CA, USA
Tris base: 2-Amino-2-(hydroxymethyl)-1,3-propanediol ($NH_2C(CH_2OH)_3$)	Sigma-Aldrich	St Louis, MO, USA
Tris hydrochloride ($NH_2C(CH_2OH)_3$) HCl	Sigma-Aldrich	St Louis, MO, USA
UltraPure™ DEPC Treated Water	Life Technologies	Carlsbad, CA, USA
Yeast tRNA, 10mg/mL	Applied Biosystems	Foster City, CA, USA

2.1.2 Commercial kits

Commercial kits used throughout this study are summarized in Table 2-2. Optimization and alteration to commercial kit protocol is given in the methods section of the associated chapter.

Table 2-2: List of commercial kits

Commercial Kit Name	Supplier	Supplier Location
Bioanalyzer RNA 6000 Nano Kit	Agilent Technologies	Santa Clara, CA, USA
dNTP Set, 100nM, PCR Grade	Life Technologies	Carlsbad, CA, USA
EZ DNA Methylation™ Gold Kit	Zymo Research	Irvine, CA, USA
GeneChip® WT PLUS Reagent Kit	Affymetrix, Inc.	Santa Clara, CA, USA
Human HCT116 DKO Methylated and Non-methylated DNA Set	Zymo Research	Irvine, CA, USA
Infinium HumanMethylation450 BeadChip Kit	Illumina, Inc.	San Diego, CA, USA
KAPA Biosystems Library Quantification Complete Kit, optimized for Roche LC480	GeneWorks Pty Ltd	Thebarton, SA, AUS
MinElute PCR Purification Kit	Qiagen	Hilden, GER
MiSeq Reagent Kit v2, 300 cycles	Illumina, Inc.	San Diego, CA, USA
NEBNext® Ultra™ Library Preparation Kit	New England BioLabs Inc.	Beverly, MA, USA
PCR Low Ladder Marker Set	Sigma-Aldrich	St Louis, MO, USA
Platinum® *Taq* DNA Polymerase, 10x buffer and $MgCl_2$	Life Technologies	Carlsbad, CA, USA
Qubit® dsDNA BR Assay Kit	Life Technologies	Carlsbad, CA, USA
Qubit® dsDNA HS Assay Kit	Life Technologies	Carlsbad, CA, USA
RNase-free DNase set	Qiagen	Hilden, GER

Commercial Kit Name	Supplier	Supplier Location
RNeasy Mini Kit	Qiagen	Hilden, GER
TruSeq DNA PCR-free LT Sample Preparation Kit, Set A	Illumina, Inc.	San Diego, CA, USA
TruSeq DNA PCR-free LT Sample Preparation Kit, Set B	Illumina, Inc.	San Diego, CA, USA
UltraClean® Tissue & Cells DNA Isolation Kit	GeneWorks Pty Ltd	Thebarton, SA, AUS
Wizard® SV Gel & PCR Clean Up System	Promega	Madison, WI, USA

2.1.3 Products and equipment

Products and equipment used for this study are summarized in Table 2-3.

Table 2-3: List of products and equipment

Product/Equipment Name	Supplier	Supplier Location
2100 Bioanalyzer	Agilent Technologies	Santa Clara, CA, USA
BioSpec-nano Spectrophotometer	Shimadzu	Kyoto, JAP
BX51 research system microscope	Olympus	Tokyo, JA
ChemiDoc™ XRS Imaging System	Bio-Rad	Hercules, CA, USA
Illumina HiSeq 2500	Illumina, Inc.	San Diego, CA, USA
LightCycler® 480 Instrument II Real-Time PCR System	Roche Diagnostics	Mannheim, GER
Magnetic stand-96	Life Technologies	Carlsbad, CA, USA
NanoDrop Spectrophotometer	ThermoFisher Scientific	Waltham, MA, USA
Polytron PT3100 rotor stator homogenizer	Kinematica	Lucerne, Switzerland
PowerPac™ Basic Power Supply	Bio-Rad	Hercules, CA, USA
Vacutainer® 10mL K_2EDTA tube	Becton, Dickinson and Company (BD)	Franklin Lakes, NJ, USA
Vacutainer® system	Becton, Dickinson and Company (BD)	Franklin Lakes, NJ, USA
Veriti 96-well ThermalCycler	Applied Biosystems	Foster City, CA, USA
Wide Mini-Sub Cell GT electrophoresis cell	Bio-Rad	Hercules, CA, USA

2.1.4 Software, genome browsers and online tools

The software, genome browsers and online tools used for this study are summarized in Table 2-4. Website links are included as Appendix 1.

Table 2-4: List of software, genome browsers and online tools

Software/Genome Browser/Tool	Supplier/Source	Supplier Location
Affymetrix® Expression Console™	Affymetrix, Inc.	Santa Clara, CA, USA
Bioconductor minfi package 3.2	Bioconductor	Seattle, WA, USA
Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7	National Institutes of Health	Berthesda, MD, USA
GenomeStudio v2011.1 with Methylation module 1.9.0	Illumina, Inc.	San Diego, CA, USA
HaplotypeCaller (Genome Analysis Toolkit (GATK) variant discovery tools)	Broad Institute	Cambridge, MA, USA
Illumina Experiment Manager (IEM) v1.9	Illumina, Inc.	San Diego, CA, USA
Integrative Genomics Viewer (IGV)	Broad Institute	Cambridge, MA, USA
KAPA Library Quantification Data Analysis Template	GeneWorks Pty Ltd	Thebarton, SA, AUS
Linear **M**odels for **M**icroarray **A**nalysis (LIMMA) package	Bioconductor	Seattle, WA, USA
MethPrimer: designing primer pairs for methylation PCRs	The Li Lab, Peking Union Medical College Hospital	Peking, China
Multiple Primer Analyzer	ThermoFisher Scientific	Waltham, MA, USA
Transcriptome Analysis Console (TAC) software	Affymetrix, Inc.	Santa Clara, CA, USA
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) assembly	University of California, Santa Cruz	Santa Cruz, USA

2.2 General methods

2.2.1 Nucleic acid isolation from esophageal and control tissues

2.2.1.1 DNA isolation

DNA was isolated from esophageal and control tissues using MO BIO's UltraClean® Tissue & Cells DNA Isolation Kit (Cat #12334-50), as per manufacturers instructions with the exception of eluting in nuclease-free water. The suggested second 50µL elution was implemented for increased yield. Briefly, 700µL of TD1 solution (thoroughly shaken) was added to each tube of dry beads. A section of thawed tissue biopsy was added directly to the beads/TD1, followed by 15µL of 20mg/mL Proteinase K. Tubes were taped horizontally to a vortex and shaken at 2800 rpm for 10 minutes, then incubated at 60°C for 30 minutes. The liquid sample was transferred to a spin filter and centrifuged at 10,000 xg for 1 minute. 400µL of TD2 was added to each column then centrifuged at 10,000 xg for 30 seconds. The column was dried by centrifugation at 10,000 xg for 1 minute. DNA was eluted in 50µL nuclease-free water, after 2 minutes incubation at room temperature by centrifugation at 10,000 xg for 30 seconds. A second elution was performed using a further 50µL nuclease-free water, incubating at room temperature 2 minutes, then centrifuging at 10,000 xg for 30 seconds. All centrifugation steps were performed at room temperature. Isolated DNA was quantified (refer to Section 2.2.1.3 Estimation of nucleic acid concentration) then stored at -80°C.

2.2.1.2 RNA isolation

RNA was isolated from fresh frozen tissue samples following Polytron PT3100 rotor stator homogenization, using the RNeasy Mini Kit (Qiagen, Cat #74104) according to the manufacturers protocol with the exception of reapplication of 30µL nuclease water for elution (following 3min incubation at room temperature). On-column DNA digestion was performed using the RNase-Free DNase Set (Qiagen, Cat #79254).

Briefly, 10µL 2-Mercaptoethanol was added per 1mL RLT buffer. Thawed sections of tissue biopsies were weighed and placed in 600µL prepared RLT buffer. Tissue was homogenized on ice using the Polytron PT3100 rotor stator homogenizer for 30 seconds at 13,000 rpm, then centrifuged at 14,000 rpm for 3 minutes. The lysate supernatant was added to an equal volume of 70% ethanol and bound to an RNeasy spin column by centrifugation at 10,000rpm for 15 seconds. RNA was washed using 350µL Buffer RW1 by centrifugation at 10,000 rpm for 15 seconds. On-column DNase digestion was performed using Qiagen's RNase-Free DNase Set by adding 10µL of prepared DNase I enzyme to 70µL RDD buffer per sample, adding to the column and incubating at room temperature for 15 minutes. RNA was washed after DNase treatment using 350µL Buffer RW1 by centrifugation at 10,000 rpm for 15 seconds. Two further wash steps using 500µL RPE buffer were performed by centrifugation at 10,000 rpm for 15 seconds (first wash) and 2 minutes (second wash). The column was dried by centrifugation at 10,000 rpm for 1 minute. RNA was eluted in 30µL nuclease-free water, after 3 minutes incubation at room temperature by centrifugation at 10,000 rpm for 1 minute. A second elution was performed by reapplying the eluate, incubating at room temperature 3 minutes, then centrifugation at 10,000 rpm for 1 minute. All centrifugation steps were performed at room temperature. Isolated RNA was quantified (refer to Section 2.2.1.3 Estimation of nucleic acid concentration) then stored at -80°C.

2.2.1.3 Estimation of nucleic acid concentration

Genomic DNA yield and estimated purity (based on absorbance ratios 260nm/280nm and 260nm/230nm), was determined using the BioSpec-Nano Spectrophotometer, using 2µL DNA, setting the instrument to double-stranded DNA (dsDNA) and pathlength to 0.7mm. Blanking was performed on nuclease-free water.

Total RNA yield, estimated purity and integrity were determined using the BioSpec-Nano Spectrophotometer and Agilent 2100 Bioanalyzer, with the Bioanalyzer RNA 6000 Nano Kit (Agilent, Cat #5067-1511). Total RNA yield and purity was estimated with the BioSpec-nano Spectrophotometer, using

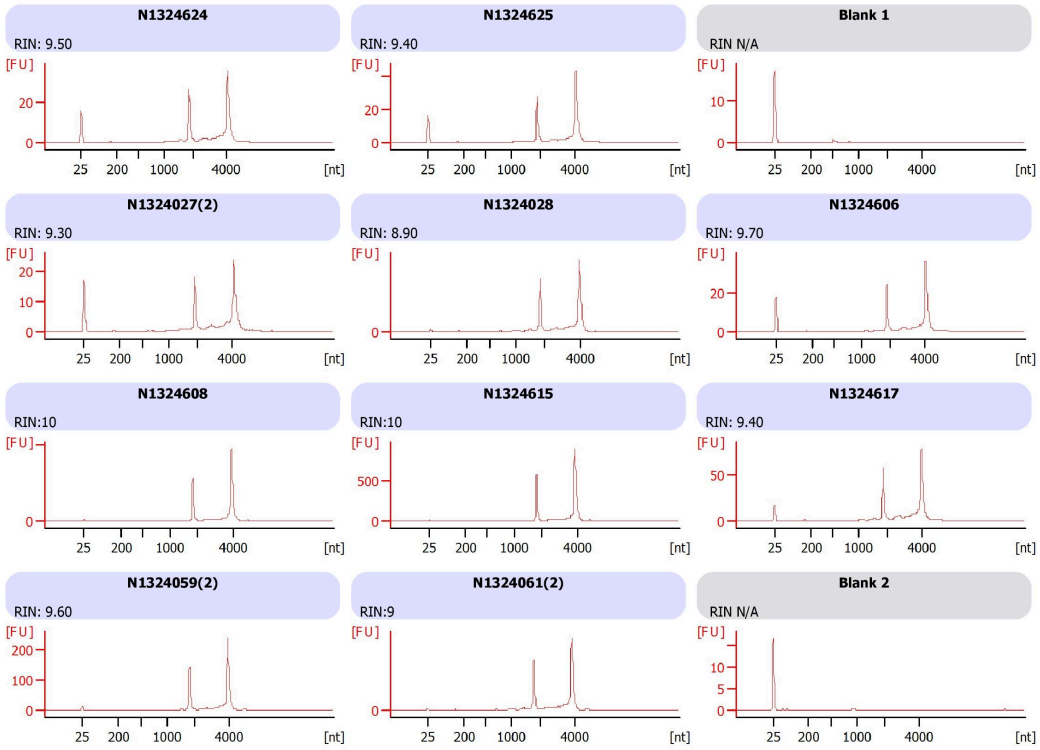
2μL of RNA, setting the instrument to RNA and pathlength to 0.7mm. Blanking was performed on nuclease-free water. RNA integrity and yield were determined using the Agilent 2100 Bioanalyzer, as outlined in Section 2.2.1.4 Determination of RNA integrity.

2.2.1.4 Determination of RNA integrity

Bioanalyzer RIN (0 – 10) was used to estimate the integrity of total RNA using electrophoretic trace and takes into account the presence or absence of degradation products. The integrity of RNA extracted from esophageal and control tissues was determined using the Agilent 2100 Bioanalyzer and Bioanalyzer RNA 6000 Nano Kit (Agilent, Cat #5067-1511) according to the manufacturer's instructions.

Briefly, all reagents were allowed to equilibrate to room temperature for 30min, protecting the dye concentrate from light. The RNA 6000 Nano dye concentrate was vortexed for 10 seconds and centrifuged briefly. 65μL of pre-filtered gel matrix was combined with 1μL of dye concentrate, vortexed thoroughly and centrifuged at 14,000 rpm for 10 minutes at room temperature. The gel-dye mix was protected from light and allowed to equilibrate to room temperature for 30 minutes prior to use. 9μL of gel-dye mix was loaded into the appropriate well of a RNA Nano chip and pressurized for 30 seconds to distribute the gel. A further 9μL of gel-dye mix was loaded into 2 additional gel-dye wells. 5μL of RNA 6000 Nano marker was added to the ladder well and each of the 12 sample wells. 1μL of thawed RNA ladder (previously heat denatured for 2 minutes at 70°C, aliquoted and stored at -80°C) was added to the ladder well. Samples were heat denatured at 70°C for 2 minutes to minimize secondary structure formation, then stored on ice. 1μL of each sample was added to each of the 12 sample wells. The chip was vortexed for 60 seconds at 2400 rpm in an IKA vortex mixer and analyzed in the Agilent 2100 Bioanalyzer within 5 minutes. Successful RNA isolation results in 2 ribosomal peaks (18S and 28S, at ~2400 and 4000nt). A marker peak at 25nt will be present for all samples.

(A)



(B)

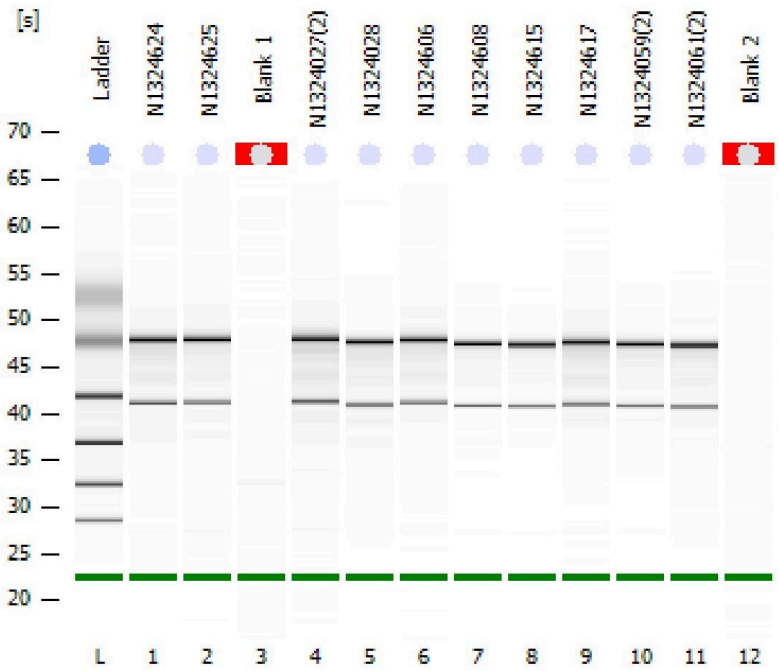


Figure 2-1: RNA integrity of 10 esophageal tissue samples analyzed using the Agilent Bioanalyzer RNA 6000 Nano Kit. Total RNA was found to be intact with minimal degradation, evidenced by RIN of 8.90 – 10. Two ribosomal peaks, 18S and 28S (visualized at approximately 2400 and 4000nt respectively), as well as the marker peak are evident for all esophageal

samples. (A) Electrophoretic traces. Sample blanks contain the marker peak only. (B) Gel separation. Esophageal samples show clean 18S and 28S bands.

2.2.2 Agarose gel electrophoresis

Agarose gels were used to resolve DNA samples such as PCR products and prepared libraries for amplicon sequencing. Briefly, gels were 2.5% agarose (in 1x Tris/Borate/EDTA (TBE) buffer), using SYBR® Safe DNA Gel Stain for visualization in 30 lanes/gel format. Amount of DNA loaded per lane varied according to expected fragment size. DNA samples were diluted in nuclease-free water and a final concentration of 1x loading buffer (Sigma Aldrich, PCR Low Ladder Marker Set, Cat #D7808). The molecular weight marker included a 20bp and 100bp ladder, loading 8µL/lane at a final concentration of 1x loading buffer, 0.4µg/µL 20bp ladder, 0.08µg/µL 100bp ladder (Sigma Aldrich, PCR Low Ladder Marker Set, Cat #D7808). The resultant ladder contained dsDNA fragments from 20 to 1000bp in 20bp increments. Gels were run at 100V (constant) for 90 minutes. Visualization was by ultraviolet transillumination performed using the Bio-Rad ChemiDoc™ XRS Imaging System.

2.2.3 DNA recovery

2.2.3.1 Wizard® SV Gel and PCR Clean Up System

Promega's Wizard® SV Gel and PCR Clean Up system (Cat #A9282), is a membrane-based purification system (binding up to 40µg DNA) for direct clean up of PCR products (100bp – 10kb). This system was used for clean up of BSP amplified target regions prior to library preparation when all pooled amplicons were > 100bp.

Briefly, an equal volume of membrane binding solution was added to the PCR amplification product, mixed, then placed onto an SV mini-column and incubated at room temperature for 1 minute then centrifuged at 16,000 xg for 1 minute. The column was washed using 700µL of membrane wash solution, centrifuged at 16,000 xg for 1minute, then repeated using a further 500µL of

membrane wash solution and increasing centrifugation time to 5 minutes. The column was re-centrifuged for 1 minute with the lid off to allow for complete evaporation of residual ethanol. Elution was performed using 50µL of nuclease-free water, incubating 1 minute at room temperature, then centrifuging at 16,000 xg for 1 minute. All centrifugation steps were performed at room temperature. Purified DNA was quantified (Section 2.2.4 Qubit® fluorometric double-stranded DNA quantification) then stored at -20°C.

2.2.3.2 Qiagen Min Elute PCR Purification Kit

Qiagen's Min Elute PCR Purification Kit (Cat #28004) is a column-based purification system (5µg maximum binding capacity) for direct clean up of PCR products (70bp – 4kb), used for clean up of BSP amplified target regions prior to library preparation when pooled amplicons include targets < 100bp. Protocol was performed as per the manufacturers instructions, with the exception of pH indicator addition, omitted due to interference with downstream sequencing; and alterations to elution procedure for increased yield.

Briefly, 5 volumes of Buffer PB were added to 1 volume of PCR reaction and passed over the column by centrifugation at 13,000 rpm for 1 minute. DNA was washed using 750µL Buffer PE by centrifugation at 13,000 rpm for 1 minute, then dried by an additional 1 minute centrifugation at maximum speed. Samples were air-dried a further 1min at room temperature with the lid open to ensure complete evaporation of all residual ethanol. DNA was eluted in 10µL nuclease-free water, incubating 1 minute at room temperature, then centrifuging 1 minute at 13,000rpm. Elution was repeated with a further 10µL nuclease-free water. Finally, 10µL of eluate was reapplied, incubated and centrifuged as before to result in ~18-19µL purified DNA. All centrifugation steps were performed at room temperature. Purified DNA was quantified (Section 2.2.4 Qubit® fluorometric double-stranded DNA quantification) then stored at -20°C.

2.2.3.3 Solid Phase Reversible Immobilization (SPRI) beads

SPRI beads are paramagnetic beads that reversibly bind DNA in the presence of polyethylene glycol (PEG) and salt. The volumetric ratio of beads to DNA can be altered to allow for size-selection manipulation. Illumina SPRI beads (Sample Purification Beads (SPB)) were used for MiSeq library preparation, more specifically for clean up of very small fragments and DNA concentration after end-repair, prior to A-tailing and again for clean up of the adapter ligated library, removing adapter dimers and other small fragments. Results from investigation of column-based versus bead-based clean up for library preparation optimization (amplicon sizes 70 – 200 bp) is covered in Section 5.3.2.1 Column versus bead-based clean up.

2.2.4 Qubit® fluorometric double-stranded DNA quantification

Qubit® fluorometric assays are highly selective for quantification of double stranded DNA (dsDNA). The Qubit® dsDNA BR Assay Kit (Life Technologies, Cat #Q32850) is a broad range kit, accurate for initial sample concentrations 100pg/μL – 1000ng/μL and was used for quantification of amplified regions of interest prior to pooling and pooled amplicons prior to library preparation. The Qubit® dsDNA HS Assay Kit (Life Technologies, Cat #Q32851) is a high sensitivity kit, accurate for initial sample concentrations 10pg/μL – 100ng/μL and was used for quantification of library adapter-ligated DNA. The single pooled library was accurately quantified for PCR-competent sequencing templates using KAPA quantification (Section 5.2.4.3 Quantification of PCR-competent sequencing template).

Qubit® assays (broad range and high sensitivity) were carried out according to the manufacturer's instruction. Briefly, all reagents were equilibrated to room temperature for 30 minutes, protected from light. A working solution was created by mixing reagent to buffer in a 1:200 ratio. Two standards were prepared, mixing 10μL of each supplied standard with 190μL of working solution. Samples were prepared by mixing 1-2μL of DNA with working solution to 200μL (based on expected sample concentration). All samples and standards were vortexed 2-3 seconds and read using the Qubit®

Fluorometer, running a new standard curve each time. Resultant reads were adjusted to account for initial dilution.

2.2.5 DNA methylation studies

2.2.5.1 DNA lysis

DNA was lysed to ensure all protein was cleaved prior to bis-conversion, aiding in achieving maximal conversion efficiency. Genomic DNA (500ng) was incubated in 1x DNA lysis buffer (2ng/µL tRNA, 280ng/µL Proteinase-K, 1% SDS), in a final volume of 20µL, for 1 hour at 37°C.

2.2.5.2 Bisulfite conversion of DNA

Bisulfite conversion of DNA was performed using Zymo's EZ DNA Methylation Gold Kit (Cat #D5005) as per the manufacturer's instructions, with the exception of alterations to the elution protocol. Briefly, 130µL of prepared CT-conversion reagent was added to 500ng lysed genomic DNA (in 20µL final volume), and heat denatured at 98°C for 10 minutes, incubated at 64°C for 2½ hours, then cooled to 4°C. 600µL of M-binding buffer was added to a Zymo spin column, followed by the converted DNA and mixed by inversion. Converted DNA was bound to the column by centrifugation at 14,000 rpm for 30 seconds, then washed using 100µL M-wash buffer and centrifuging at 14,000 rpm for 30 seconds. Desulfonation was performed by incubation with 200µL M-desulfonation buffer for 15 minutes at room temperature, and centrifuging at 14,000 rpm for 30 seconds. After desulfonation, DNA was washed twice using 200µL M-wash buffer and centrifuging at 14,000 rpm for 30 seconds. The column was air-dried for 2 minutes with the lid open to allow for complete evaporation of residual ethanol. Bis-DNA was eluted following a 2 minute incubation in 10µL M-elution buffer and centrifuging at 14,000 rpm for 30 seconds. The eluate was re-applied to the column to increase yield, incubated 2 minutes and centrifuged at 14,000 rpm for 30 seconds. Following this, a further 10uL nuclease-free water (Qiagen nuclease-free water, Cat #129114) was added to the column, incubated for 2 minutes, then eluted by centrifugation for 30

seconds at 14,000 rpm (total eluate volume ~18-19uL). All centrifugation steps were performed at room temperature.

2.2.5.3 Quantification of bisulfite converted DNA

The concentration of bis-converted DNA can be quickly estimated using a spectrophotometer. Bisulfite-converted DNA is single stranded with limited non-specific base pairing at room temperature, thus its absorption coefficient at 260nm resembles that of RNA. Therefore, the concentration of bisulfite-converted DNA can be estimated using the Shimadzu BioSpec Nano Spectrophotometer, setting the instrument to RNA and pathlength to 0.7mm (using a value of 40ug/mL for Ab260 = 1.0).

2.2.5.4 Bisulfite-specific PCR (BSP)

Bisulfite-specific PCR was used to amplify bis-converted regions of interest for targeted amplicon sequencing. Oligonucleotide primers, specific to the bisulfite-modified sequence were designed to facilitate this analysis. Assays were optimized to ensure unbiased amplification of methylated and unmethylated DNA input. Sections 5.2.3.1 and 5.3.1.2 outline the primer design process and resultant BSP assays.

Duplicate (optimization) or triplicate (technical and independent validation cohorts) PCR amplification was performed for each region of interest using primers as described in Table 5-6. Amplicon genomic location is summarized in Table 5-5. BSP assays were performed in a final volume of 10µL, containing 5-10ng bisulfite-converted DNA template, 200µM dNTPs, 200nM each primer, 1x buffer without $MgCl_2$, 1.0 – 1.5mM $MgCl_2$ and 0.033U/µL Platinum® *Taq* DNA polymerase. SYTO® 9 was used for optimization of BSP assays (1:20 dilution) but not included for amplification prior to targeted sequencing due to possible interference. SYTO® 9 (excitation maximum: 483nm, florescence emission maximum: 503nm) is less inhibitory to PCR than the traditionally used SYBR Green I and results in highly reproducible, robust melt curves[79].

Amplification was performed using the Roche LightCycler ® 480 Instrument II Real-Time PCR system, in 96-well plate format (Roche LightCycler ® 480 Multiwell Plate 96, Cat #04 729 692 001) under the following conditions: 95°C, 2 min x 1 cycle; 95°C, 40 sec / χ°C, 90 sec / 72°C, 2 min x 5 cycles; 95°C, 40 sec / χ°C, 1 min / 72°C, 90 sec x 35-40 cycles; 72°C, 5 min x 1 cycle, cool to 40°C, hold 10 sec; where χ represents the optimized annealing temperature for the specific BSP assay. For melting curve analysis (BSP optimization only) the following melt program was added prior to cooling: 95°C, 15 sec x 1 cycle; 65°C, 1 min x 1 cycle, continuous acquisition (5 acquisitions/°C) to 97°C. Ramp speeds were set to maximum (4.4°C/s for heating, 2.2°C/s for cooling). Table 5-7 outlines the optimal assay conditions for each primer set.

2.2.5.5 DNA methylation detection

Heat dissociation melt curve analysis was used to determine bias in amplification for BSP assay optimization (Section 2.2.5.4 outlines SYTO® 9 addition and melt curve cycling). Melt curve analysis (T_m calling) was performed using the Roche LightCycler ® 480 Instrument II Real-Time PCR system. Initially, bisulfite-converted Human HCT116 DKO Methylated (M) and Non-Methylated (U) DNA as well as Human Genomic DNA from blood (B) was used for optimization (Figure 2-2). Later, as Human Genomic DNA proved to be entirely unmethylated in all regions of interest, it was adopted as the unmethylated control (U) and CpGenome Universally Methylated DNA was used as a fully methylated (M) control. BSP assays were optimized to ensure a 50:50 mix of bisulfite-converted unmethylated/methylated control amplified in proportion to amplification of fully methylated and fully unmethylated controls.

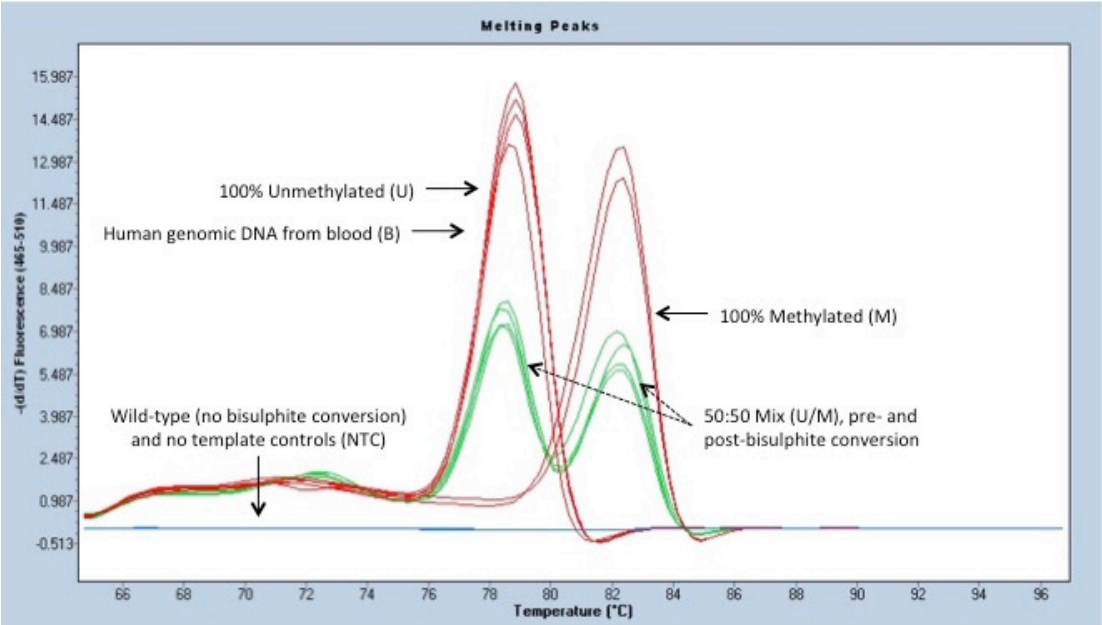


Figure 2-2: Determination of DNA methylation status using heat dissociation melt curve analysis. Bisulphite-specific PCR amplification of Chr19:44,455,288-44,455,448, 160bp, containing 11 CpG sites. Fully methylated DNA dissociates at 82.2+/-0.2°C, fully unmethylated DNA dissociates at 78.6+/-0.2°C. Duplicate inputs of fully methylated bisulphite-converted DNA (M), fully unmethylated bisulphite converted DNA (U), bisulphite converted commercially available human genomic DNA from blood (B), 50:50 mix of U and M, both pre- and post-bisulphite conversion, wild-type control (non-bisulphite converted 50:50 mix of U and M) and no template control (NTC) were run. The 50:50 mixture of methylated and unmethylated DNA input (both pre- and post-bisulphite conversion) show unbiased amplification, in proportion to U and M controls, using 2mM $MgCl_2$ and 58°C annealing. For this target region, human genomic DNA from normal blood (B) is entirely unmethylated, important for a disease-associated hypermethylation blood biomarker.

CHAPTER 3: PATIENT COHORT AND SAMPLE SELECTION

3.1 Introduction

Despite detection and reproducibility problems, histopathological evaluation of esophageal tissue biopsies is still the clinical 'gold standard' for the surveillance and treatment of Barrett's esophagus and its dysplastic progression[80]. As outlined in Chapter 1, this combined endoscopic and histologic diagnosis method has a number of pitfalls, leading to problematic patient classification. During endoscopy, abnormality must be correctly identified and then accurately sampled. The Seattle Protocol, recommended by guidelines worldwide[52, 81], is based on the premise that systematic, high-count sampling will minimize the possibility of missing 'worst case' diagnosis. Further confounding accurate diagnoses are technical issues with preparation of H&E tissue sections and inherent biological heterogeneity of tissue biopsy, further increasing the possibility of misclassification[43].

Intra-observer and even inter-observer variability in dysplasia diagnosis, in particular low-grade dysplasia, is a problem recognized almost 30 years ago[82], that still exists today, even among highly experienced specialist gastrointestinal pathologists[52, 83, 84].This may be due in part to inadequate tissue sections (poor preparation as well as biopsy sampling error) or atypia related to inflammation or ulceration. However, even with perfect sampling and technical preparation of H&E sections, dysplasia assessment remains highly variable. The risk of progression from non-dysplastic BE to EAC is relatively low, with a rate of ~0.2-0.4% per year[85], however progression from low to high-grade dysplasia or EAC may be as high as 85% at 9 years[86]. Thus, detection of LGD is really the most important stage for identification of treatable, early stage disease; but clinically is the most difficult and often eludes detection. The low-grade dysplasia conundrum: the need for accurate detection, yet the most problematic identification, has important implications for research studies that rely on correct classification of sample and patient. Misclassification will impact downstream analysis and quality of result. Numerous diagnostic biomarkers have been proposed for BE and EAC[52, 87],

many of which provide contradictory evidence for clinical utility, perhaps due in part to this conundrum.

When selecting patient cohorts and samples for biomarker development studies, attention must be paid to variation in attributes such as patient demographic, treatment stage and history, as well as source of non-tumor specimen and selection of appropriate controls. Matching attributes such as age, sex, race and life-style characteristics such as smoking habits between disease and control groups is important for the reliable identification of potentially relevant biomarkers[78].

3.1.1 Ethical considerations

This study falls under ethics approval already obtained for the Australian-wide research group "Progression of Barrett's Esophagus to Cancer Network" (PROBE-NET). PROBE-NET encompasses a large number of studies Australia-wide, supporting research into understanding biological mechanisms for the progression of BE to EAC, identification and testing of biomarkers, developing strategies to prevent or delay disease progression and developing better treatments and novel therapies.

More specifically, this study comes under the umbrella of biomarker identification and testing, as outlined on p27 of the PROBE-NET NEAF: "***Our primary aim is to identify those factors associated with progression from normal esophageal mucosa to reflux esophagitis to BE to EAC*** *across the full range of potential modifiers,* ***including biomarkers (such as levels of RNA or protein expression, or methylation****), phenotype (including BMI), medical conditions (such as gastroesophageal reflux, type II diabetes), environmental factors (smoking, alcohol) and pharmacologic factors (in particular, anti−acid medications, aspirin and non-steroidal anti−inflammatory drugs [NSAIDs])."*

Ethical and scientific approval from the Human Research Ethics Committee (HREC) and site-specific approval (SSA) for this project was renewed in early 2014. HREC reference HREC/13/SVH/344, granted 14th January 2014,

approves human tissue and blood collection at St Vincent's Hospital, Reginald V N Lord's private rooms (St Vincent's Clinic, Suite 606, Level 6), Westmead hospital and Nepean hospital. Site specific approval for St Vincent's Hospital required renewal during the project, and was approved on the 19th February 2014 (valid until 14th January 2019), SSA reference SSA/14/SVH/46. Copies of the confirmation letters (HREC/13/SVH/344, SSA/14/SVH/46) are included as Appendix 2.

3.1.2 Chapter 3 aims

To apply stringent criteria to classify and select suitable samples for methylation and expression profiling, targeted amplicon sequencing and pan-cancer mutation screening.

3.2 Methods

3.2.1 Biospecimen collection and processing

Human esophageal, duodenal and proximal stomach tissue biopsies and peripheral blood were obtained from consenting patients undergoing standard endoscopy or esophagectomy, prior to sedation.

3.2.1.1 Tissue biopsy collection and processing

Tissue samples were placed directly into RNA*later*® RNA Stabilization Reagent (Life Technologies, Cat #AM7020) to stabilize and protect RNA, eliminating the need for immediate processing or freezing. In brief, endoscopic biopsies ≤ 4mm were placed into 400µL RNA*later*® RNA Stabilization Reagent, whilst larger surgical samples (> 5mm) were cut into smaller pieces and placed into 1mL of RNA*later*®. Samples were stored at 4°C for a minimum of 24 hours to allow full permeation of the preservative solution into the sample.

A central section was taken of each biopsy, formalin-fixed, embedded in paraffin, and stained with hematoxylin and eosin (H&E). For surgical samples with multiple tissue fragments, a section was taken for each fragment, placed in a separate tube with sample number maintained and suffixed with a

unique identifier, for example: N1427142(2). Each section was evaluated by a minimum of two independent pathologists, blinded to patient details and hospital diagnosis. Tissue samples were stored at -20°C for 1 month, then residual RNA*later* removed and moved to permanent -80°C storage. Standard 'non-research' routine biopsies were taken as part of all endoscopic procedures and analyzed independently by hospital pathologists to determine patient diagnosis.

3.2.1.2 Blood collection and processing

Blood was drawn from consenting patients, prior to sedation, via cannula into BD Vacutainer® 10mL K_2EDTA tubes (Cat #367525), up to 40mL per patient. Immediately after collection, each tube was mixed by inversion and transported to the laboratory at room temperature for processing. Whenever possible, blood processing was completed within 8hrs from time of collection, avoiding possible contamination of cell-free circulating DNA (cfc-DNA) with DNA released from leukocytes lysed during longer pre-processing times[88]. A double-spin protocol was utilized, comprising initial 10 minute centrifugation at 3000rpm, separating plasma (transferred to RCF 30,000 xg rated tubes), from buffy coat and red blood cell layers (discarded). A subsequent 10 minute centrifugation of the isolated plasma was performed at 16,000 xg to enrich for cell-free, or exosome-associated DNA (fragments of nucleic acid not associated with cells or cell fragments will remain in suspension at this speed, whereas cellular-associated DNA is pelleted out and discarded). Plasma supernatant containing cfc-DNA was transferred to cryogenic vials (Sigma-Aldrich, Nunc® CryoTubes® 1.8mL internal thread, Cat #V7634-500EA), without disturbing the pellet (visible for >1mL of plasma), and stored at -80°C. All centrifugation steps were performed at 4°C.

3.2.2 Data collection and storage

Patient and sample data were recorded on standardized forms at each visit from a consenting patient. On consent, a patient ID was assigned and remained unchanged for return visits. Patient ID stickers were affixed to every page of the sample and patient data forms.

Patient and sample data were collated into password-protected, searchable databases, recording fields outlined in Table 3-1. Hardcopy originals of patient and data collection forms were scanned and stored in pdf format.

Table 3-1: Database fields recorded for patient and biospecimen tracking. General data were recorded (patient and sample) at the time of collection, and associated biospecimen nucleic acid extraction data updated as performed. §: calculated automatically.

	Patient-specific fields	**Tissue-specific fields**	**Blood-specific fields**
General data	Patient ID, Surname, First name, Sex, DOB, Visit #, Endoscopic diagnosis, Pathology-confirmed diagnosis, Reflux, Reflux surgery, Prague criteria, Procedure comments, Hospital pathology.	Tissue sample ID, Date collected, Age at collection§, Pathologist sample evaluation, Distance (cm), Site, Appearance, Tissue storage location, Comments.	Blood sample ID, Date collected, Age at collection§, EDTA tube count, Collection time, Time processing complete, Processing time§, Plasma volume (mL), Storage location, Comments.
DNA isolation	N/A	Tissue DNA code, Extraction date, Protocol used, Lysis time, DNA concentration (ng/µL), 260/280, 260/230, DNA volume (µL), Total DNA (µg), Storage location.	Plasma DNA code, Extraction date, Protocol used, Extraction format, Volume plasma used (mL), Volume plasma remaining§, Thaw time (min), Thaw until lysis time (min), Lysis temperature (°C), Time through column (min), DNA concentration (ng/µL), DNA volume (µL), Total DNA (ng), Storage location, Extraction notes.
RNA isolation	N/A	Tissue RNA code, Extraction date, Protocol used, Tissue weight (mg), RNA	N/A

	Patient-specific fields	Tissue-specific fields	Blood-specific fields
		concentration (ng/μL), 260/280, 260/230, RIN, RNA volume (μL), Total RNA (μg), Storage location.	

3.2.3 Patient data collection

As part of patient consent, additional data (in the form of a health questionnaire) were collected (enrollment only; not return visits). The health questionnaire records the following items:

- Frequency of reflux
- Frequency of heartburn
- Medications (previous and current)
- Previous operations
- Family history (reflux, heartburn, Barrett's esophagus, esophageal cancer)
- Current weight (kg), height (cm) and waist circumference (cm); as well as most ever weighed (kg)
- Smoking history
- Alcohol consumption history
- Previous medical conditions (diabetes (and how it was treated), high blood pressure, stomach cancer, esophageal cancer, other cancers)

The patient explicitly consents (or can choose to disallow) researchers to access their medical, oncology and pathology records.

3.2.4 Sample data collection

Sample data were recorded on standardized forms at the time of collection. Duplicate sample number stickers (one for the tissue or blood tube, the other for the sample collection form) ensure the possibility of error was minimized.

The sample collection form consists of a:

- General data sheet
- Tracking log form
- Clinical classification form
- Tissue biopsy report and blood collection form
- Histology classification form

3.2.4.1 General data sheet

The general data sheet recorded patient ID, visit number, patient name, address, telephone number and email, GP name, Medicare number and two contact person details. The invitation date (date the patient was invited to the study) was also recorded; generally via post (letter and information pack sent) in the weeks prior to their procedure, or on the day (for return visits). An endoscopy clinic or hospital patient sticker was affixed to this page.

Procedure information (procedure date, endoscopist, hospital, doctor and confirmation that written consent was obtained) was also recorded on the general data sheet.

3.2.4.2 Tracking log form

A log form was used for biospecimen tracking and to ensure all documentation was complete. Dates were recorded (date issued/requested, date received/completed, and date to Queensland Institute of Medical Research (QIMR, centralized database for all biospecimen collection under the ProbeNet umbrella)), for each of the eight documentation components:

1. Information brochure
2. Consent form
3. Health questionnaire
4. Endoscopy report
5. Clinical classification form
6. Pathology report
7. Biopsy report form
8. Histology classification form

For biospecimen collection, the log form recorded the date collected, number of sample tubes, storage location and date stored for tissue and blood samples.

3.2.4.3 Clinical classification form

A clinical classification form was completed during the procedure and included sections for recording:

- Primary indication for the procedure (reflux, dyspepsia, BE surveillance, treatments, pre- / post-surgical, etc…)
- Distance to the gastroesophageal junction (GOJ) aboral (cm)
- Clinical diagnosis for the patient (no abnormality detected, reflux esophagitis, BE, EAC, adenocarcinoma of the cardia, squamous cell carcinoma (SCC), etc…)

The recorded clinical classification was signed and dated by the endoscopist or doctor performing the procedure.

3.2.4.4 Tissue biopsy report and blood collection form

The procedure date, site and doctor were recorded on this form. An endoscopy clinic or hospital patient sticker was affixed to this page.

Collected biospecimen data included:

- Tissue biopsy data:
    - Tissue sample number (sticker affixed to form, duplicate sticker affixed to sample tube). A unique sample number was used for each biopsy.
    - Distance from the GOJ (cm)
    - Biopsy site (esophageal, upper stomach, duodenal, other)
    - Biopsy appearance (columnar, squamous, tumor, etc…)
    - Any comments
- Blood data:
    - Blood sample number (sticker affixed to form, duplicate sticker affixed to blood tube). The same sample number was applied to

ALL tubes collected of a specific type (EDTA tubes for plasma). If serum was also collected (for other studies), a new sample number was applied to ALL serum tubes.

- Time of collection
- Type of blood collection tube (EDTA, serum, etc…)
- Any comments

3.2.4.5 Histology classification form

The histology classification form was completed based on *hospital* patient classification (worst diagnosis detected in routine 'non-research' biopsies, analyzed independently by hospital pathologists to determine patient diagnosis). Classification was signed and dated by the endoscopist or doctor responsible for the patient.

3.2.5 Patient and sample classification

Patient classification was based on hospital biopsy sampling and pathology (as recorded on the histology classification form (Section 3.2.4.5)). Research ***sample classification*** was based on blinded, independent evaluation of H&E stained slides from a representative section of sample tissue, performed by a minimum of two pathologists (as outlined in Section 3.2.1.1). For identification of Barrett's esophagus (BE), the presence of goblet cell containing columnar mucosa was required.

3.2.6 Patient selection for study inclusion

In all cohorts, patients were age and gender matched (see Table 3-2: Patient cohort summary, p51), with patients free from any other forms of cancer (both presently and previously).

3.2.7 Sample selection for study inclusion

Stringent quality requirements were imposed on extracted nucleic acid to ensure successful downstream methylation and expression profiling, targeted amplicon sequencing and pan-cancer mutation screening. Yields of greater than 500ng and 1.5µg were required for tissue RNA and DNA respectively.

Absorbance ratios for 260nm/280nm were required to fall in the range 1.70 – 2.00 and 260nm/230nm: 1.70 – 2.30. RNA integrity number (RIN; RNA only) was required to be ≥ 7.50 to proceed with expression profiling.

Tissue samples for the study were classified as normal squamous (N), BE without dysplasia (BE), BE with low-grade dysplasia (LGD), BE with high-grade dysplasia (HGD) or esophageal adenocarcinoma (EAC). Samples classified as "indefinite for dysplasia" were excluded. Uniformity in clinical treatment regimes and pathological definition allowed for grouping of HGD and EAC classifications for data analysis purposes. Minimum 60% homogeneity was required for inclusion of individual tissue biopsies in the study, however, in practice, the majority of biopsies selected (with the exception of LGD) were pathology-confirmed 80-100% homogeneity (Table 3-3).

3.2.8 Nucleic acid isolation from esophageal and control tissues

DNA and RNA were isolated as per Section 2.2.1 from each selected tissue biopsy that met inclusion criteria described in Section 3.2.7 Sample selection for study inclusion. A number of extraction kits were trialed and optimized to ensure optimal yield and quality of nucleic acid for downstream applications (results outlined in Section 3.3.4 Optimization of nucleic acid isolation from esophageal tissue).

3.3 Results

3.3.1 Patient and biospecimen collections

Over the course of the study (February 2013 – February 2016), 3058 tissue and 566 blood samples were collected from 319 unique patients (127 normal, 70 non-dysplastic BE, 1 BE-ID (indefinite for dysplasia), 17 LGD, 22 HGD, 27 EAC and 55 'Other'). 'Other' includes classifications such as esophagitis, cardiac mucosa, and squamous cell carcinoma. Repeat visits comprised 43.64% of all visits (247 in 566), breakdown as outlined in Figure 3-1. Patients consenting to medical record access enabled full previous histories (prior to enrolment in the study) to be extracted.

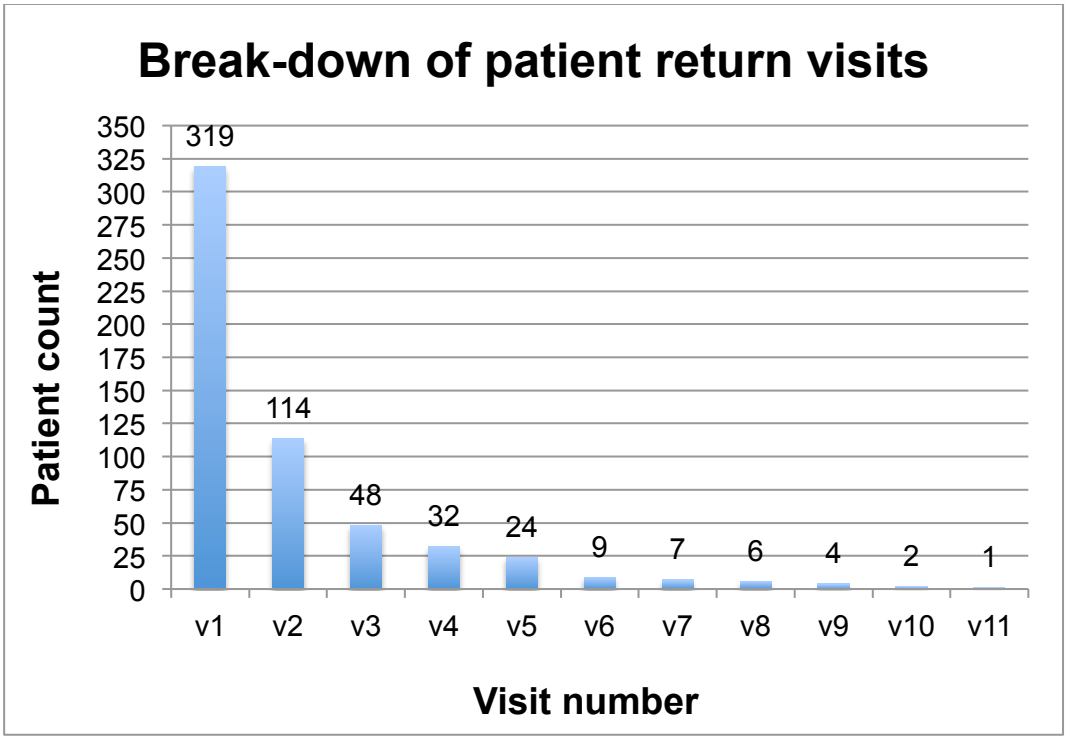


Figure 3-1: Patient return visits over the duration of the study. 319 unique patients enrolled in the study over 3 years with patients returning up to 10 times for follow-up and treatment.

Nucleic acid was extracted from 437 esophageal and control tissue samples (245 DNA and 185 RNA) for downstream applications such as genome-wide methylation and expression profiling, targeted sequencing and pan-cancer mutation screening (not including samples for protocol optimization, Figure 3-2). Of the 245 tissue DNA extractions, associated classification breakdown is as follows: 9 Duodenal, 9 Proximal stomach, 123 N, 43 non-dysplastic BE, 3 BE-ID, 16 LGD, 19 HGD, 16 EAC and 7 'Other', including esophagitis and cardiac mucosa samples. Of the 192 tissue RNA extractions, associated classification breakdown is as follows: 9 Duodenal, 9 Proximal stomach, 99 N, 26 non-dysplastic BE, 3 BE-ID, 9 LGD, 17 HGD, 13 EAC and 7 'Other', including esophagitis and cardiac mucosa samples. Note here that classification refers to *sample* classification, not patient diagnosis (matched normal proximal samples from diseased patients were extracted and analyzed).

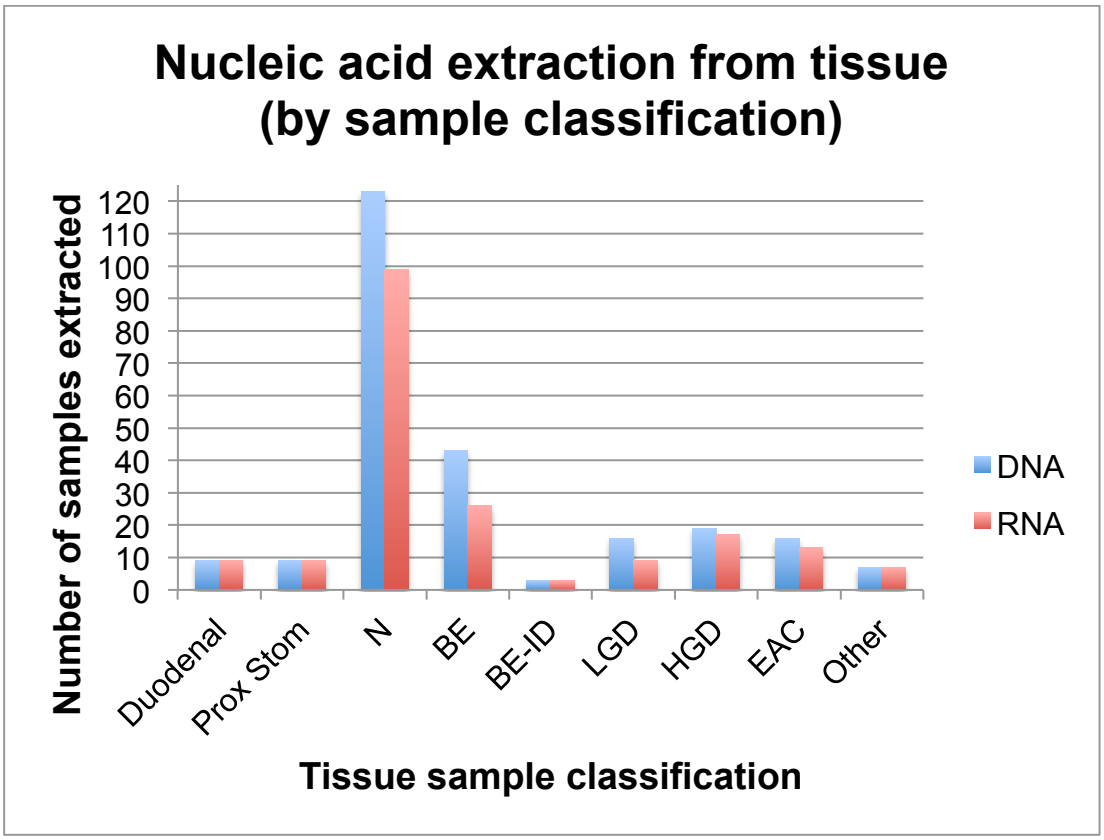


Figure 3-2: Nucleic acid extraction from esophageal and control tissues by sample classification. Tissue samples were selected for nucleic acid extraction based on pathologist evaluation of the associated H&E stained section (≥ 60% homogeneity). Resultant nucleic acid yield and purity, along with sample homogeneity and patient characteristics were aspects for consideration when selecting samples for downstream applications.

Cell-free circulating DNA was extracted from 210 plasma samples and stored at -80°C for anticipated blood biomarker investigative studies. Of these 210 plasma samples, the associated classification breakdown is as follows: 71 N, 43 non-dysplastic BE, 3 BE-ID, 10 LGD, 22 HGD, 15 EAC and 46 'Other'; where 'Other' refers to patients suffering from squamous cell carcinoma, esophagitis, cardiac mucosa and esophageal inflammation. Note here that classification of blood samples is synonymous with patient diagnosis.

3.3.2 Patient cohorts

This study used five patient cohorts (summarized in Table 3-2): a training cohort, technical validation cohort, independent validation cohort, the Cancer Genome Atlas cohort and an external validation cohort.

Genome-wide methylation (HM450) and expression (HTA2.0) profiling was performed on a training cohort (n=48 samples from 27 patients). A subset of this (technical validation cohort, n=24 samples from 18 patients) underwent targeted amplicon sequencing (10 target regions) to validate observed disease-associated differential methylation. An independent validation cohort (n=24 samples from 15 patients) independently confirmed these results, examining the previous ten, plus an additional eight (total 18) regions of interest. A large EAC-focused external validation cohort from collaborators in Queensland[49] was also used to validate differential methylation findings.

3.3.2.1 Training cohort

Genome-wide methylation and expression profiling was performed on a training cohort of 48 samples from 27 patients, divided into two subsets. The first subset consisted of DNA and RNA extracted from 24 esophageal tissue samples taken from 12 patients (3 patients from each of the four classifications: normal esophagus (N), Barrett's esophagus (BE), Barrett's esophagus with low-grade dysplasia (LGD) and combined Barrett's esophagus with high grade dysplasia and esophageal adenocarcinoma (HGD/EAC)). A proximal (25cm, matched normal) and distal sample (reflecting disease classification, or in the case of a normal patient, normal distal esophagus tissue) was included for each selected patient.

The second subset of the training cohort consisted of 24 samples from 15 patients (5 patients each from the three classifications: BE, LGD, HGD/EAC). A distal sample was taken from each of the 15 patients (reflecting disease classification), with a matched normal, duodenal and proximal stomach sample taken from a representative case from each of the disease classifications.

The first subset confirmed no difference in molecular signature from matched normal samples (taken proximally from diseased patients) to that from normal healthy patients, irrespectively of whether taken proximally or distally (results Section 4.3.2.1 Normal squamous tissue methylation) and tight sample clustering, indicative of minimal patient-to-patient difference. The second subset was profiled subsequently to analysis of the first and thus more focused on diseased samples, where molecular signature was found to be more variable.

3.3.2.2 Technical validation cohort

The technical validation cohort consisted of a subset of the training cohort, used to validate selected target regions differentially methylated in genome-wide profiling. Targeted amplicon sequencing was used for validation. The cohort comprised 24 samples from 18 patients, selected primarily on tissue homogeneity.

3.3.2.3 Independent validation cohort

The independent validation cohort was used to independently validate selected target regions differentially methylated in genome-wide profiling using targeted amplicon sequencing. This cohort consisted of 24 samples from 15 patients. Two control samples, CpGenome universally methylated DNA (Millipore, Cat #S7821) and fully unmethylated DNA (Human genomic DNA from blood, Roche, Cat #11 691 112 001) were included in this cohort. The cohort also contained a time-series of multiple samples from a single patient to examine methylation changes with disease treatment.

3.3.2.4 External validation cohort

An external validation cohort, consisting of genome-wide methylation profiling data (Illumina Infinium HumanMethylation450 BeadChip®) from a total of 250 samples from 154 patients was obtained with the kind permission of A/Prof. Andrew Barbour from the University of Queensland, Translational Research Institute at the Princess Alexandra Hospital. This data formed a part of a

comprehensive characterization of methylation in esophageal adenocarcinoma, identifying a CIMP subtype and was published in February 2016[49]. The external validation cohort was comprised of 125 EAC, 19 BE (11 adjacent BE from EAC patients, 8 BE samples from BE patients), 85 non-tumor squamous esophagus (10 GERD, 75 normal squamous) and 21 normal stomach samples. No intermediate stages of dysplasia were analyzed.

Table 3-2: Patient cohort summary. Four cohorts were used, a Training cohort, Technical validation cohort, Independent validation cohort and External validation cohort. The external validation cohort also consisted of 10 samples (disease-affected distal esophagus) taken from control patients suffering from GERD, average age 55.7 ± 10.8 years, 60.0% male, 40.0% female (data not shown). Average Prague Criteria rounded to the nearest whole integer.

	Total (n)	Duodenal	Proximal Stomach	N	BE	LGD	HGD	EAC
Training Cohort								
Patients (n)	27	N/A	N/A	15	9	7	3	5
Samples (n)	48	3	3	18	9	7	3	5
Average age ± Standard deviation (yrs)		68.0 ± 15.0	68.0 ± 15.0	68.2 ± 10.8	63.2 ± 11.2	66.9 ± 6.9	69.0 ± 12.1	70.4 ± 8.6
Gender: Male (%)		66.7	66.7	83.3	77.8	71.4	100	80.0
Female (%)		33.3	33.3	16.7	22.2	28.6	0	20.0
Average Prague Criteria (Cx Mx)		N/A	N/A	N/A	C1 M3	C3 M6	C2 M4	C2 M2
Technical Validation Cohort								
Patients (n)	18	N/A	N/A	3	6	3	2	4
Samples (n)	24	0	0	9	6	3	2	4
Average age ± Standard deviation (yrs)		N/A	N/A	69.0 ± 12.4	62.0 ± 10.0	65.0 ± 9.0	73.0 ± 14.8	71.0 ± 9.9
Gender: Male (%)		N/A	N/A	88.9	83.3	100.0	100.0	75.0
Female (%)		N/A	N/A	11.1	16.7	0.0	0.0	25.0
Average Prague Criteria (Cx Mx)		N/A	N/A	N/A	C1 M3	C4 M8	C2 M4	C3 M3

	Total (n)	Duodenal	Proximal Stomach	N	BE	LGD	HGD	EAC
Independent Validation Cohort								
Patients (n)[1]	15	N/A	N/A	0	6	4	3	4
Samples (n)[2]	24	0	0	5	6	4	3	4
Average age ± Standard deviation (yrs)		N/A	N/A	62.0 ± 9.5	59.0 ± 11.1	667.0 ± 13.0	62.0 ± 14.0	59.0 ± 7.0
Gender: Male (%)		N/A	N/A	100.0	83.3	100.0	66.7	100.0
Female (%)		N/A	N/A	0.0	16.7	0.0	33.3	0.0
Average Prague Criteria (Cx Mx)		N/A	N/A	N/A	C3 M5	C9 M10	C7 M8	C5 M8
External Validation Cohort								
Patients (n)	154	0	N/A	11	8	0	0	125
Samples (n)	250	0	21	75	19	0	0	125
Average age ± Standard deviation (yrs)		N/A	63.6 ± 11.8	61.5 ± 12.3	66.6 ± 10.2	N/A	N/A	62.6 ± 10.5
Gender: Male (%)		N/A	90.5	87.5	86.7	N/A	N/A	92.5
Female (%)		N/A	9.5	12.5	13.3	N/A	N/A	7.5

[1] This cohort contained a patient with consecutive follow up samples to examine methylation changes with disease treatment (HGD, LGD and BE sample taken as patient classification improved)

[2] Two control samples (fully methylated and fully unmethylated standard material) were included in the independent cohort (details not shown)

3.3.3 Sample cohorts

Nucleic acid yield and purity, sample homogeneity and patient characteristics were aspects for consideration when selecting samples for downstream applications. Sample quality characteristics of each cohort are summarized in Table 3-3. Typically, higher nucleic acid yield was obtained from samples at more advanced stages of disease; however the highest yields were recorded for control duodenal and proximal stomach samples.

Absorbance ratios A260/280 and A260/230 reflect sample purity, with pure DNA resulting in A260/280 of approximately 1.8; RNA ~2.0. Pure nucleic acid will have A260/230 ~2.0 – 2.2. Ratios are affected by nucleotide mix as well as pH and ionic strength of nucleic acid eluate; therefore were used as a guide only. Overall, DNA cohorts showed reproducible A260/280 ~1.90; RNA ~2.10-2.15; concordant with contaminant-free nucleic acid. A260/230 was much more variable: DNA ~ 1.85-2.30; RNA 1.70-2.20.

A minimum of 60% homogeneity was required for inclusion of individual tissue biopsies in the study, however, in practice, the majority of biopsies selected (with the exception of LGD) were pathology-confirmed 80-100% homogeneity (Table 3-3).

Table 3-3: Sample quality within each patient cohort. All characteristics are mean ± standard deviation. Homogeneity averages performed over all available pathologist evaluation. Technical and independent validation cohorts used for targeted DNA sequencing only. Sample quality control data not available for the External Validation cohort. NA: nucleic acid, RIN: RNA integrity number.

	Total NA (µg)	RIN	A260/280	A260/230	Homogeneity (%)
DNA cohorts					
Training Cohort					
Duodenal	16.39 ± 3.17	-	1.91 ± 0.01	2.19 ± 0.06	100.0 ± 0.0
Proximal stomach	9.51 ± 3.62	-	1.92 ± 0.01	2.16 ± 0.08	100.0 ± 0.0
Normal	4.37 ± 1.65	-	1.90 ± 0.06	1.89 ± 0.24	100.0 ± 0.0

	Total NA (µg)	RIN	A260/280	A260/230	Homogeneity (%)
Non-dysplastic BE	3.14 ± 1.58	-	1.91 ± 0.11	1.97 ± 0.20	91.6 ± 13.6
LGD	7.53 ± 6.61	-	1.89 ± 0.05	1.97 ± 0.18	66.3 ± 14.8
HGD-EAC	9.84 ± 6.90	-	1.90 ± 0.03	2.12 ± 0.14	81.9 ± 33.5
Technical Validation Cohort					
Normal	4.08 ± 1.86	-	1.88 ± 0.08	1.98 ± 0.30	100.0 ± 0.0
Non-dysplastic BE	3.50 ± 1.96	-	1.87 ± 0.08	1.99 ± 0.17	95.6 ± 7.2
LGD	10.89 ± 10.77	-	1.86 ± 0.02	1.87 ± 0.14	78.1 ± 3.9
HGD-EAC	11.98 ± 6.67	-	1.88 ± 0.01	2.12 ± 0.17	96.7 ± 6.1
Independent Validation Cohort					
Normal	3.83 ± 1.10	-	1.87 ± 0.10	2.14 ± 0.44	100.0 ± 0.0
Non-dysplastic BE	5.94 ± 2.12	-	1.89 ± 0.08	2.33 ± 0.13	98.9 ± 3.0
LGD	7.29 ± 2.48	-	1.87 ± 0.06	2.28 ± 0.35	63.8 ± 4.0
HGD-EAC	8.60 ± 3.12	-	1.89 ± 0.06	2.19 ± 0.36	92.5 ± 9.0
RNA cohorts					
Training Cohort					
Duodenal	5.78 ± 1.53	8.27 ± 0.75	2.15 ± 0.05	2.21 ± 0.07	100.0 ± 0.0
Proximal stomach	10.38 ± 5.77	8.23 ± 0.81	2.14 ± 0.02	2.19 ± 0.02	100.0 ± 0.0
Normal	1.46 ± 1.08	9.08 ± 0.53	2.16 ± 0.07	1.71 ± 0.19	100.0 ± 0.0
Non-dysplastic BE	3.44 ± 1.36	9.11 ± 0.60	2.10 ± 0.02	1.82 ± 0.18	91.6 ± 13.6
LGD	4.72 ± 5.00	9.10 ± 0.63	2.12 ± 0.02	1.88 ± 0.27	66.3 ± 14.8
HGD-EAC	3.97 ± 2.86	8.93 ± 0.80	2.13 ± 0.01	1.86 ± 0.26	81.9 ± 33.5

3.3.4 Optimization of nucleic acid isolation from esophageal tissue

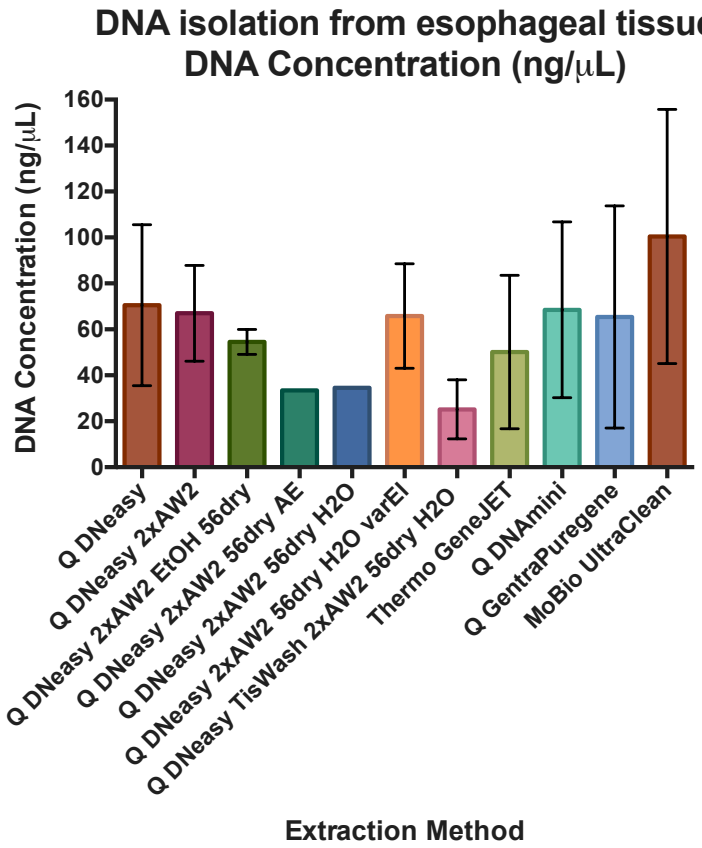
A number of extraction kits were trialed and optimized to ensure optimal yield and quality of nucleic acid for downstream applications.

3.3.4.1 DNA isolation

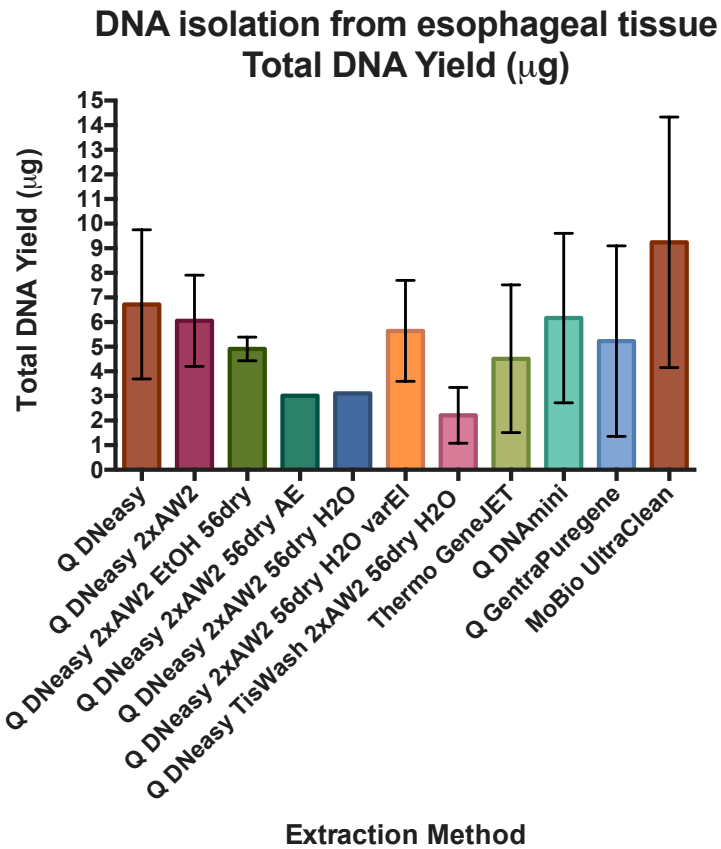
Five alternative kits for DNA isolation were trialed and optimized (Table 3-4), with the best yield and purity attained using Mo Bio's UltraClean® Tissue & Cells DNA Isolation Kit (Figure 3-3). The most significant problems encountered were with low 260nm/230nm absorbance ratios, indicative of phenol, guanidine or other organic contamination. We hypothesized this was due to high salt concentration from the RNA*later*, which fully penetrates the tissue and thus was not alleviated with additional wash steps.

Table 3-4: Optimization of DNA isolation from fresh frozen esophageal tissue samples, preserved in RNA*later*. Five kits were trialed using protocol variation to optimize yield and purity. Due to small biopsy size, different tissue biopsies were used for the comparison, however they were all of the same tissue type (normal squamous tissue) and equivalent input amounts used.

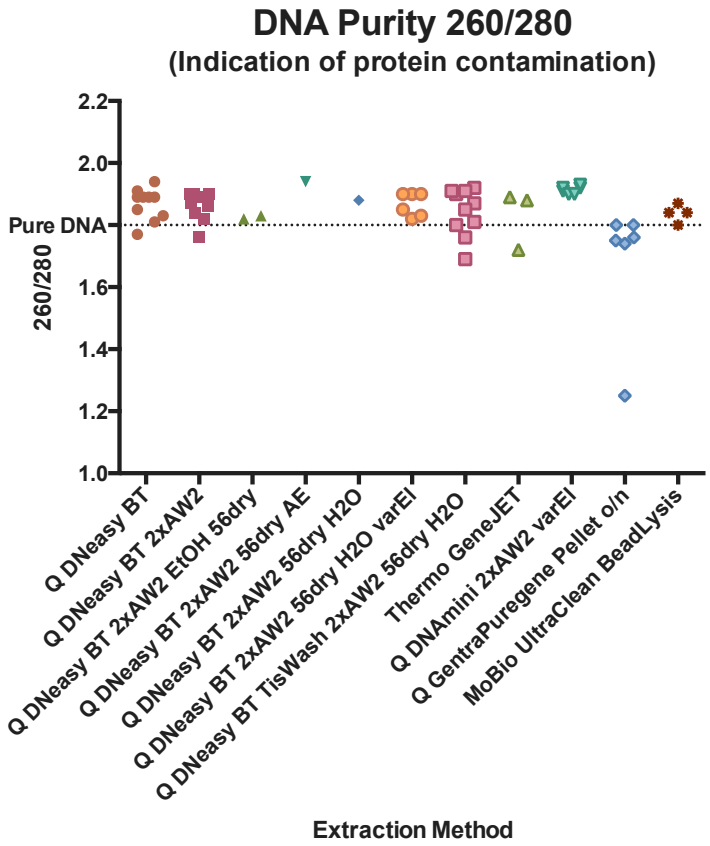
Supplier	Kit	Protocol Variation	Samples tested
Qiagen	DNeasy Blood & Tissue Kit	Reapply eluate	10
		2x AW2 wash	10
		2x AW2 wash, 2x EtOH wash, 56°C 10min dry	2
		2x AW2 wash, 56°C 10min dry, AE successive elution	1
		2x AW2 wash, 56°C 10min dry, H_2O successive elution	1
		2x AW2 wash, 56°C 5min dry, variable elution	6
		Tissue wash, 2x AW2 wash, 56°C 5min dry	10
Qiagen	QIAamp DNA Mini Kit	Variable elution (reapply v successive)	6
Qiagen	Gentra Puregene Tissue	Overnight rehydration	6
Thermo Fisher	GeneJET Genomic DNA Purification	2 x 50µL elution	3
MO BIO	UltraClean® Tissue & Cells DNA Isolation	Elution in H_2O rather than TD3	4


DNA isolation from esophageal tissue
DNA Concentration (ng/μL)
DNA Concentration (ng/μL)
160
140
120
100
80
60
40
20
0
Q DNeasy
Q DNeasy 2xAW2
Q DNeasy 2xAW2 EtOH 56dry
Q DNeasy 2xAW2 56dry AE
Q DNeasy 2xAW2 56dry H2O
Q DNeasy 2xAW2 56dry H2O varEl
Q DNeasy 2xAW2 56dry H2O
Q DNeasy TisWash 2xAW2 56dry H2O
Thermo GeneJET
Q DNAmini
Q GentraPuregene
MoBio UltraClean
Extraction Method

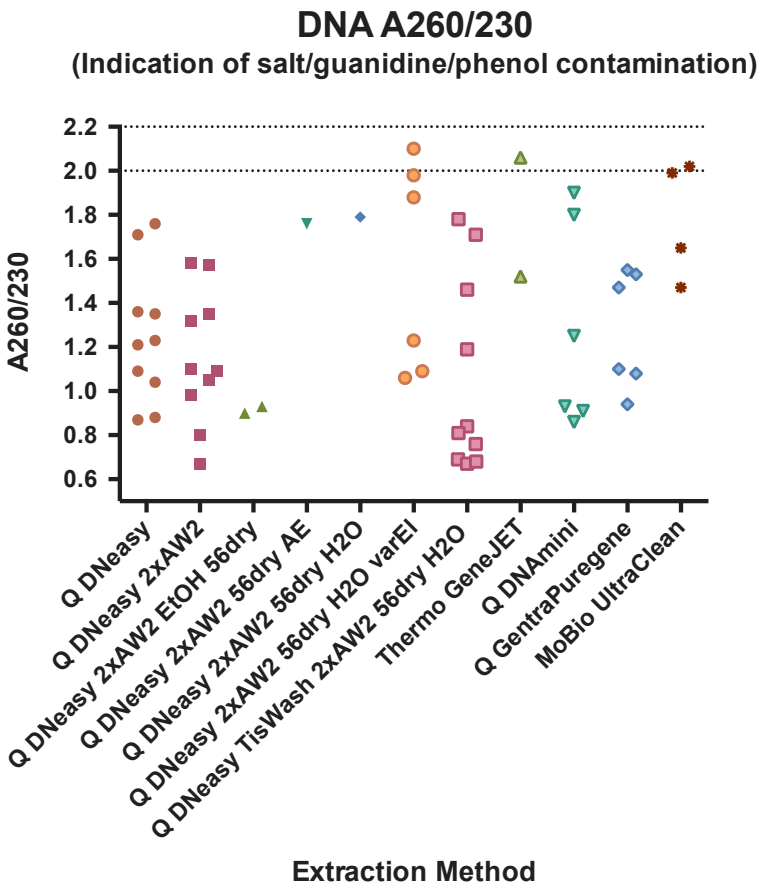
(A)


DNA isolation from esophageal tissue
Total DNA Yield (μg)
Total DNA Yield (μg)
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Q DNeasy
Q DNeasy 2xAW2
Q DNeasy 2xAW2 EtOH 56dry
Q DNeasy 2xAW2 56dry AE
Q DNeasy 2xAW2 56dry H2O
Q DNeasy 2xAW2 56dry H2O varEl
Q DNeasy 2xAW2 56dry H2O
Q DNeasy TisWash 2xAW2 56dry H2O
Thermo GeneJET
Q DNAmini
Q GentraPuregene
MoBio UltraClean
Extraction Method

(B)


DNA Purity 260/280
(Indication of protein contamination)
260/280
Pure DNA
2.2
2.0
1.6
1.4
1.2
1.0
Q DNeasy BT
Q DNeasy BT 2xAW2
Q DNeasy BT 2xAW2 EtOH 56dry
Q DNeasy BT 2xAW2 56dry AE
Q DNeasy BT 2xAW2 56dry H2O
Q DNeasy BT 2xAW2 56dry H2O varEl
Q DNeasy BT 2xAW2 56dry H2O
Q DNeasy BT TisWash 2xAW2 56dry H2O
Thermo GeneJET
Q DNAmini 2xAW2 varEl
Q GentraPuregene Pellet o/n
MoBio UltraClean BeadLysis
Extraction Method

(C)


DNA A260/230
(Indication of salt/guanidine/phenol contamination)
A260/230
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
Q DNeasy
Q DNeasy 2xAW2
Q DNeasy 2xAW2 EtOH 56dry
Q DNeasy 2xAW2 56dry AE
Q DNeasy 2xAW2 56dry H2O
Q DNeasy 2xAW2 56dry H2O varEl
Q DNeasy 2xAW2 56dry H2O
Q DNeasy TisWash 2xAW2 56dry H2O
Thermo GeneJET
Q DNAmini
Q GentraPuregene
MoBio UltraClean
Extraction Method

(D)

Figure 3-3: Yield and purity of DNA isolation kit optimization. (A) & (B) The best yields were obtained from Mo Bio's UltraClean® Tissue & Cells DNA Isolation Kit, averaging 9.24 ± 5.09µg total DNA yield at a concentration of 100.4 ± 55.34ng/µL. (C) A260/280 ratios were acceptable for the majority of samples across all kits and protocols tested. Pure DNA will have A260/280 ~1.8; the UltraClean® kit averaged 260/280 1.838 ± 0.029. (D) A260/230 ratios were poor for the majority of samples. Best results were obtained for Mo Bio's UltraClean® (average 1.783 ± 0.268) and Thermo Fisher Scientific's GeneJET kits (all samples 260/230>1.45). Pure DNA will have A260/230 range: ~2.0-2.2.

There is some evidence of correlation between DNA yield (total DNA (µg)) and purity (260/230) (Figure 3-4). The poorest 260/230 ratios (<1.0) occur when total DNA yield is less than 5µg. This supports the observation that the best choice of kit is the one returning the highest yields (Mo Bio's UltraClean® kit).

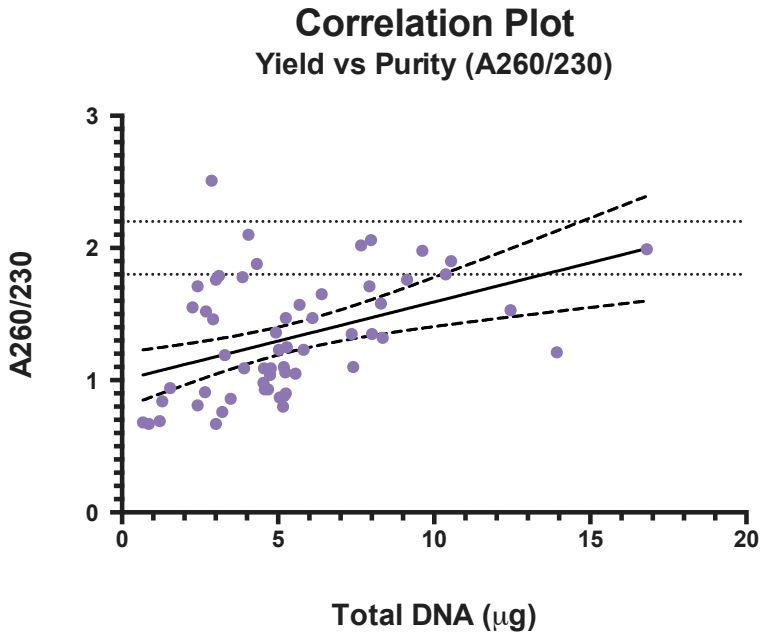


Figure 3-4: Correlation between DNA yield (µg) and purity (A260/230). Linear regression line of best fit is plotted with 95% confidence band. Pearson correlation coefficient r^2 = 0.1804 (assumes Gaussian distribution).

As a result of kit testing and optimization, DNA for this study was isolated from esophageal and control tissues using Mo Bio's UltraClean® Tissue & Cells DNA Isolation Kit (Cat #12334-50), as per the manufacturers instruction with the exception of eluting in nuclease-free water. The suggested second 50µL elution was implemented for increased yield.

3.3.4.2 RNA isolation

Four alternative kits for RNA isolation were trialed and optimized, with the best yield and purity attained using Qiagen's RNeasy Mini Kit. Yield, purity, correlation and biases in RNA isolation from esophageal tissue were examined following extraction from 154 tissue samples (Appendix 3: Yield, purity, correlation and bias in RNA isolation from esophageal tissue). Data includes trial kit extractions (Mo Bio's UltraClean® Tissue & Cells RNA Isolation Kit, Bioline's Isolate II RNA Mini Kit and Zymo's Quick-RNA^{TM} MiniPrep kit) as well as the selected RNeasy Mini Kit (Qiagen); chosen on the basis of superior absorbance ratios and RNA integrity number (RIN). As observed for DNA isolation, the most significant problem encountered was low A260/230 absorbance ratio, indicative of phenol, guanidine or other organic contamination. Again, we hypothesized this was due to high salt concentration from RNA*later*, which fully penetrates the tissue and thus is not alleviated with additional wash steps.

As a result of kit testing and optimization, RNA for this study was isolated from frozen tissue samples performing Polytron PT3100 rotor stator homogenization and using the RNeasy Mini Kit (Qiagen, Cat #74104) according to the manufacturers protocol with the exception of reapplication of 30uL nuclease water for elution (following 3min incubation at room temperature). On-column DNA digestion was performed using the RNase-Free DNase Set (Qiagen, Cat #79254).

3.4 Discussion

Despite the abundance of biomarker discovery studies for Barrett's esophagus and esophageal adenocarcinoma, none have been implemented for clinical use to date[40]. This study aims to use only exceptionally well-classified samples for genome-wide profiling as well as targeted sequencing, using publicly available and external cohorts (containing much larger sample numbers) for further validation. The study falls under the umbrella of the nation-wide collaborative research group PROBE-NET, and thus patient consent, as well as patient data and sample collection protocol was already established. For the most part, this was adopted without change, however some alterations were made, such as blood processing protocols to enrich for cancer-associated cfc-DNA.

Interestingly, on the basis of research sample pathology, a number of patients were brought in for repeat endoscopy. For example, low-grade dysplasia was detected in a particular research sample by a single pathologist and independently classified as 'indefinite for dysplasia' by another. Hospital biopsies reported non-dysplastic Barrett's only for this patient. Based on our observations, the patient was invited for repeat endoscopy and with thorough sampling, LGD classification was confirmed by hospital pathology. This highlights the problematic aspect of endoscopic sampling for patient diagnosis and the need for alternative methods.

Initial analysis of genome-wide methylation data was performed using the first subset of the training cohort (24 samples from 12 patients), prior to sample selection for the second subset. The purpose of this was primarily analyses of methylation signature of normal esophagus, however comparison of differential methylation in BE versus (LGD/HGD/EAC) was also performed to determine if our analyses had the potential to identify clinically useful methylation biomarkers. This preliminary dataset identified a number of disease-associated hyper-methylated regions. Top hits of DCBLD2, HS3ST3B1, ZNF878, ZNF844 were identified using cut-off criteria of baseline (BE) methylation < 0.10 and delta methylation (LGD and HGD

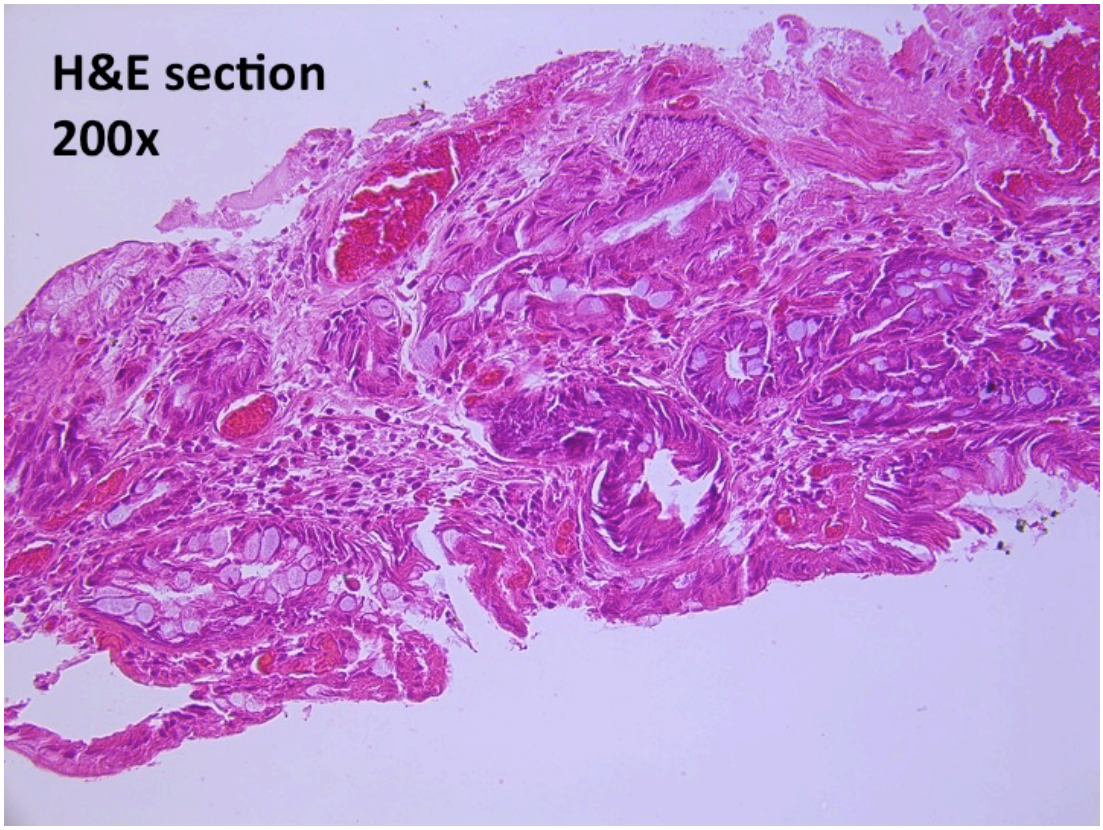
and EAC) > 0.25; and imposing criteria: N tissue and normal peripheral blood methylation < 0.10. These targets looked promising. Promoter hypermethylation inducing epigenetic down-regulation of discoidin, CUB and LCCL domain containing 2 (DCBLD2) had been previously observed in gastric cancer cell proliferation and invasion[13], but was novel for dysplastic BE and EAC. Notable in our analyses was the abundance of aberrantly methylated zinc finger loci. Hypermethylation of zinc finger protein genes has long been associated with carcinogenesis in various forms of human cancer, such as ZNF382 in acute myeloid leukemia[89], however identified loci, encoding zinc finger protein 878 and 844, had not previously been associated with EAC development.

However, using the second subset (24 samples from 15 patients) as validation for the first revealed an intriguing observation. The top 32 genes identified as differentially methylated in LGD, HGD and EAC, compared to N and BE in the first subset, showed a trend of hypermethylation in *BE*, LGD, HGD and EAC (with only minimal methylation detected in normal and control tissues and normal peripheral blood). How did this change in methylation status of non-dysplastic BE arise? Aspects such as gender bias, differences in patient age, Prague criteria, biopsy distance from GOJ, time between biopsy sampling and nucleic acid extraction, DNA concentration/purity and time from nucleic acid extraction until genome-wide profiling were analyzed: all showed no significant difference between the two subsets (data not shown). Thus it was hypothesized sample homogeneity could be responsible for this phenomenon. In view of this, a further two expert pathologists were invited to classify each of the samples involved, scoring each sample and providing an estimate composition of the representative section. In addition to this, the original two pathologists repeated their analyses of the same samples (in a blinded manner, unaware they had previously analyzed these sections). Estimated sample homogeneity between the first and second subset using the now more comprehensive sample homogeneity evaluation was analyzed and found to significantly different (Unpaired t-test, assuming Gaussian distribution, $p = 0.0093$). The problem stemmed from inclusion of LGD patients in the comparative analysis. These samples had lower

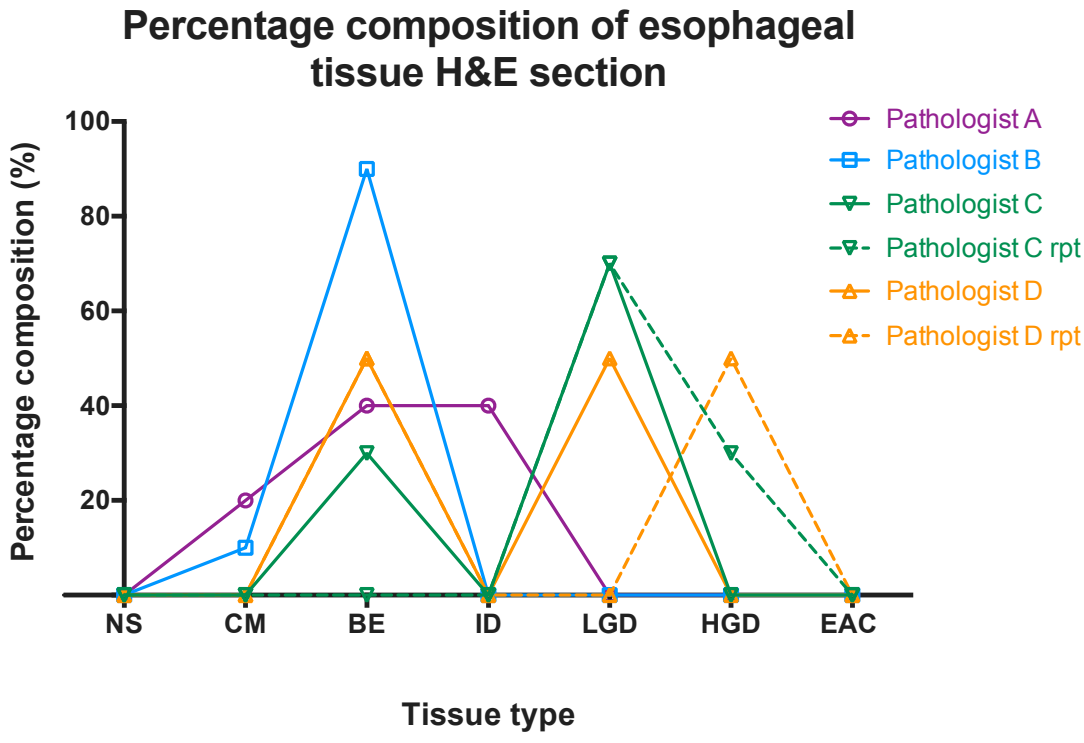
homogeneity (a 60% minimum was required for inclusion; the remainder of these biopsies tended to be non-dysplastic BE), thus skewing the results (Figure 3-5). Overall, sample composition reanalysis showed strong inter- and intra-observer agreement for classification of N, non-dysplastic BE, HGD and EAC samples, however significant variability was reported surrounding samples containing LGD.

Sample classification in the first subset had been based on patient confirmed diagnosis (thereby including LGD samples with 25%, 14% and 65% LGD (and the remainder generally non-dysplastic BE) in the LGD classification. In contrast, sample classification in the second subset was based on sample confirmed diagnosis, and included significantly fewer of these controversial 'mixed classification' samples. The lead candidates selected in subset 1, based on BE (unmethylated) vs LGD-HGD-EAC (hypermethylated) actually had some percentage of non-dysplastic BE tissue contributing to the hypermethylated LGD/HGD/EAC profile. Hence these targets, when re-examined using subset 2 data, actually showed hypermethylation in some non-dysplastic BE samples (Figure 3-5 (C)). This preliminary analysis triggered (i) the recruitment of two further pathologists (four in total) for evaluation of all difficult samples (variable reports of sample homogeneity) for all sample cohorts included in this study, (ii) the strict homogeneity requirement cut-offs implemented for the study, (iii) use of sample confirmed diagnosis, rather than patient diagnosis for classification and (iv) exclusion of controversial, difficult to diagnose LGD samples from analyses for the BE vs HGD-EAC comparison (results for LGD still examined and included).

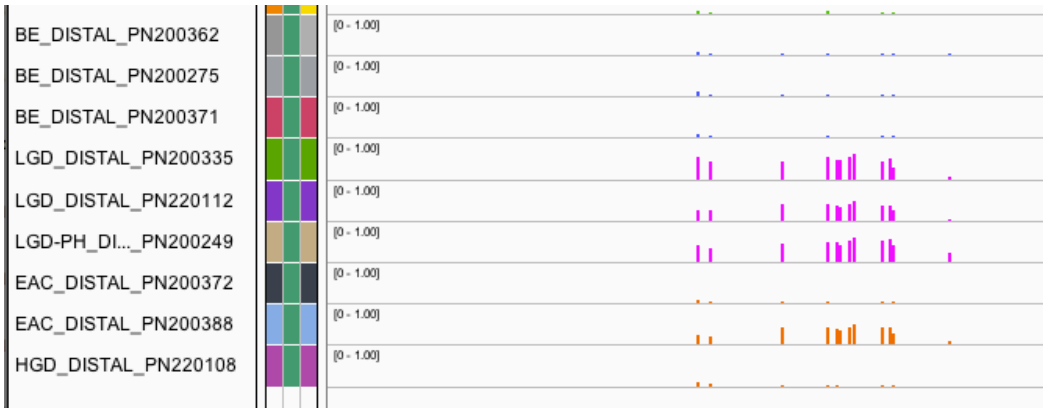
(A)


H&E section
200x

(B)


Percentage composition of esophageal tissue H&E section
Percentage composition (%)
100
80
60
40
20
NS
CM
BE
ID
LGD
HGD
EAC
Tissue type
Pathologist A
Pathologist B
Pathologist C
Pathologist C rpt
Pathologist D
Pathologist D rpt

(C) (i) DCBLD2 Subset 1



(ii) DCBLD2 Subset 2

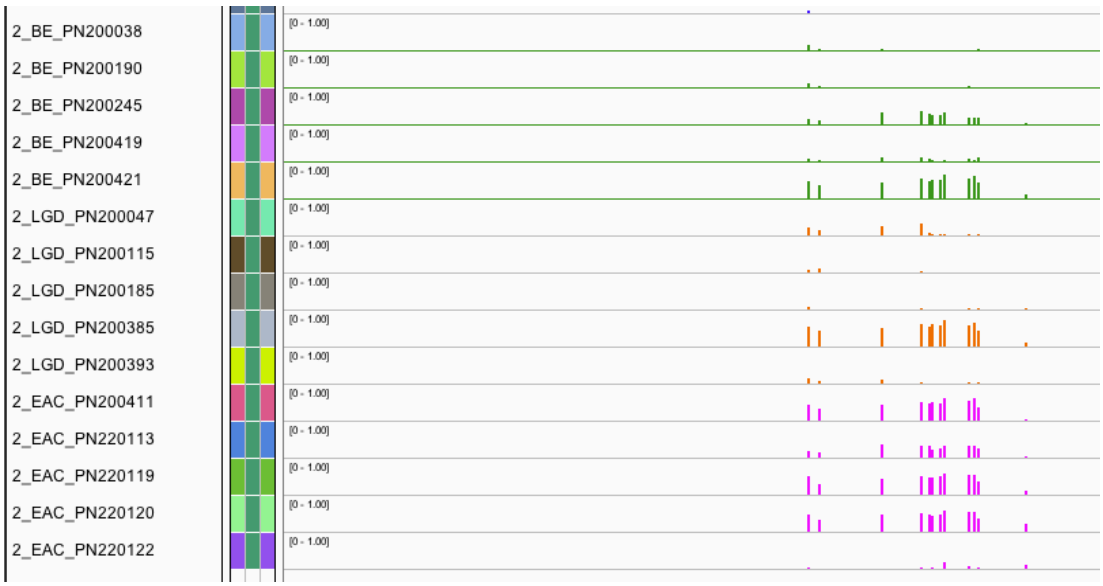


Figure 3-5: Variation in sample classification from multiple pathologists. (A) H&E stained section (200x magnification) of an esophageal tissue sample (34cm distally) taken from an 83 year old male (B) Percentage composition classification from four independent, blinded pathologists. Pathologist C and D unknowingly evaluated the same sample a second time, a minimum of 5 weeks later (sample not recognized). NS: normal squamous, CM: cardiac mucosa, BE: Barrett's esophagus, ID: indefinite for dysplasia, LGD: low-grade dysplasia, HGD: high-grade dysplasia, EAC: esophageal adenocarcinoma. (C) Impact of incorrect classification: (i) discoidin, CUB and LCCL domain containing 2 (DCBLD2), selected as a candidate for discrimination of dysplasia/EAC from non-dysplastic BE in subset 1 was

based on LGD samples which included a proportion of non-dysplastic BE tissue. (ii) This resulted in the apparent hypermethylation of this target region in non-dysplastic BE in subsequent subset 2 analyses.

The typical approach for cancer biomarker development studies is to use the current clinical 'gold standard' diagnosis method to determine sample/patient classification and evaluate potential biomarker performance with respect to this[40, 78]. As reported in literature[80], my study confirmed the classification of non-dysplastic BE, HGD and EAC to be generally undisputed, however LGD was frequently not detected or missed and showed poor reproducibility when it was detected. Many studies to date have completely avoided the inclusion of LGD samples and patients due to their inherent difficulty. However, these samples are highly valuable for identification of early-phase, treatable disease. I considered a 'reverse' approach for sample classification, using resultant molecular signature from genome-wide analyses to group similar samples post-analysis. However, with complexities such as high rates of chromothripsis[90], and a stem cell/inflammation origin rather than mutation-driven like many other cancers (Section 4.3.7 Methylation of tumor suppressor p63); this approach unnecessarily complicated sample classification and could possibly introduce further bias when deciding grouping for comparative analysis, so was discarded. My approach was to exclude LGD from all comparative groups (therefore these patients and samples did not play a role in determining disease-specific differential methylation or expression targets), but include LGD samples in all cohorts to determine if identified disease-specific changes are detectable at this early stage of carcinogenesis.

CHAPTER 4: GENOME-WIDE METHYLATION AND EXPRESSION PROFILING

4.1 Introduction

4.1.1 Current field of play

Despite the plethora of studies investigating specific methylation or expression changes in esophageal adenocarcinoma, very few studies have combined genome-wide methylation and expression profiling to characterize the role of methylation in EAC development. Krause et al[49] were the first to do this comprehensively, using Illumina's Infinium HumanMethylation450 BeadChip® (HM450) technology for methylation profiling (n=250; 125 EAC, 19 BE, 85 N, 21 Stomach) and Illumina Human HT12 (v4) for expression profiling (n=70; 48 EAC, 4 BE, 18 N), publishing their results in 2016. The group kindly shared their HM450 data for validation purposes for this study (GSE72874), prior to making the data publicly available in April 2016.

The absence of dysplastic or any intermediate development phase samples is a notable omission from the Krause et al study. Low and high-grade dysplasia are important phases in the metaplasia-dysplasia-adenocarcinoma development sequence and cannot be ignored if a full understanding of the role of methylation in EAC carcinogenesis is to be elucidated. This study is, to date, novel in its analysis of the *full* metaplasia-dysplasia-adenocarcinoma progression sequence (including both low and high grade dysplastic mucosa) using comprehensive, genome-wide methylation and expression profiling techniques. Inclusion of intermediate phases of disease development allows for elucidation of methylation and expression differences observed not only between tumor and normal (as is the case for most studies), but also between non-dysplastic precursor disease and early-phase EAC development, important for identification of potential early stage biomarkers. We anticipate differential methylation and expression for each of the comparison groups to be informative about biological changes occurring

during disease progression and possibly uncover novel changes occurring during early carcinogenesis.

4.1.2 Epigenetic involvement in EAC carcinogenesis

Genes frequently reported as hyper-methylated in BE and EAC include (but are not limited to) AKAP12, APC, CDH13, DAPK1, GPX and GST, HPP1, MGMT, NELL1, p16/CDKN2A, Reprimo (RPRM), PRDM2, RUNX3, SFRP, SOCS, SST, TAC1, TERT, TIMP3, WIF1[38, 47, 48, 50, 53, 56-63, 76, 91-93].

Epigenetic involvement in carcinogenesis is still not yet fully understood and recent publications have highlighted that DNA methylation at promoter regions of tumor suppressor genes causing transcriptional silencing / loss of methylation at oncogene promoters resulting in increased expression may only be a small part of the story. Interestingly, genome-wide methylation analysis has revealed that global hypomethylation (and not hypermethylation) is the dominant change in development of Barrett's metaplasia, but progression from BE to EAC is characterised by a smaller wave of selective promoter hypermethylation[48]. Hypermethylation sites in BE and EAC are mainly CpG-rich promoters, with the most hyper-methylated tumors showing the worst patient survival[49].

4.1.3 Methylation and expression: The bigger picture

In a genome-wide methylation and transcriptomic profiling study of normal, BE and EAC samples, Krause et al (2016)[49] reported 63% of detected differential methylation showing significant correlation with mRNA expression levels. This raises the question: what is happening with the other approximately 40%?

In a recent (2015) article examining changes in correlation between promoter methylation and gene expression in cancer, Moarii et al[94] challenged the simplistic model of direct inhibition of gene expression by promoter methylation in cancer development. They performed a large scale meta-analysis of methylation profiles of normal and cancer tissues and the

variations in expression of associated target genes and found the interplay between promoter methylation and gene expression in cancer is not simple. They found that cancer-specific methylation does not always repress gene expression but instead targets genes that are lowly expressed in normal tissues. Their findings suggest that epigenetic reprogramming may contribute to carcinogenesis in part by modifying gene expression susceptibility to changes in DNA methylation rather than a direct influence. This recent study highlights that interdependence of genetic and epigenetic alterations is poorly understood and not a simple association.

4.1.4 Pathways in EAC tumorigenesis

In the development of EAC, both genetic and epigenetic mechanisms contribute to alteration of signaling pathways that are believed to drive tumorigenesis[39]. Pathways associated with aberrant promoter methylation in BE and EAC include enrichment of known cancer signaling pathways: cell adhesion, regulation of epithelial-to-mesenchymal transition (EMT) and transforming growth factor (TGF) and WNT signaling[49]. The WNT gene family encode for secreted signaling proteins, and are involved in several developmental processes regulating cell fate as well as being implicated in carcinogenesis. Vascular endothelial growth factors (VEGFs) are involved in angiogenesis, the formation of new blood vessels, needed for tumor growth and metastatic spread of tumors. There is evidence that VEGF pathways may be important in neoplastic progression of BE to EAC[95-97].

4.1.5 The origin of BE and tumor suppressor p63 methylation

The origin of Barrett's esophagus has been a subject of considerable controversy without achieving consensus in the field[86]. Many of the common cancers show early genetic and epigenetic changes that result in cancer development; however for a subset of inflammation-induced cancers, the story is more interesting. The appearance of Barrett's esophagus and progression to EAC is similar to that of gastric IM and the progression to gastric adenocarcinoma in many ways. Both progress via a metaplasia-dysplasia-adenocarcinoma sequence with IM triggered by chronic

inflammation due to gastroesophageal reflux disease in the case of BE, and *Helicobactor pylori* infection in the case of gastric IM[98].

Initially, Barrett's esophagus was thought to originate from gastric cardia epithelium, in a migratory process to repair reflux-damaged esophageal epithelium, however the dominant concept is now centered around esophageal squamous stem cells, which when damaged by acid reflux are induced to switch their fate to the generation of columnar epithelium with intestinal characteristics[99]. The prevailing concept currently accepted in the field for the origin of BE is that of a discrete population of residual embryonic cells (RECs) that exist at the gastroesophageal junction (GEJ) in normal individuals. These RECs expand and colonize regions of the esophagus affected by chronic reflux, forming intestinal metaplasia (IM) within days of esophageal injury[86]. There is more evidence of inflammation-induced stem cell oncogenesis than mutation-driven oncogenesis for EAC; with 25% of putative Barrett's stem cells showing no cancer-related genomic aberration in a 2015 study by Yamamoto et al[100].

In 2008, Ben-Porath et al established an embryonic stem (ES) cell gene-set enrichment signature, consisting of four groups (ES expressed, NOS targets, polycomb targets, Myc targets) of 13 partially overlapping gene sets that were associated with human ES cell identity. They examined a number of aggressive human tumors and found that poorly differentiated breast, glioblastoma, bladder carcinomas display an ES-like signature, with striking correlation between tumor grade and presence of the signature, however were unable to determine whether this signature is inherited from a stem-cell-of-origin or is re-activated during tumor progression[101].

Tumor suppressor p63 is a member of the p53 transcription factor family, involved in the self-renewal of epithelial stem cells. Interestingly, it is almost never mutated in cancer, yet alterations in p63 expression are associated with tumorigenesis and chemoresistance in a number of different cancers[102]. The reduced expression of p63 in bladder carcinoma has been shown to correlate with tumor stage and grade (p63 was easily detectable in normal

bladder tissue)[103]. Park et al. also showed that abnormal expression of two major isoforms of p63 (TAp63 and ΔNp63) were observed on treatment of human bladder carcinoma cell lines with 5-Aza-2'-deoxycytidine, suggesting aberrant promoter methylation of both these isoforms in bladder carcinogenesis[103]. The residual embryonic stem cell theory for the origin of Barrett's esophagus is supported by p63 null mouse models, which quickly develop Barrett's-like tissue. On analysis, this tissue is a near-exact match with human Barrett's at the gene expression level[104]. The potential therapeutic opportunities are immense: therapies targeting RECs (which have different gene expression signatures to normal tissues) could provide an avenue for permanent eradication of BE before it progresses to cancer.

4.1.6 Chapter 4 aims

*First, to use genome-wide methylation and transcriptomic profiling to define global methylation and expression change occurring at **all** stages of esophageal adenocarcinoma development; including novel intermediate stages of Barrett's esophagus with low- and high-grade dysplasia.*
Second, to investigate DNA methylation correlation with expression of neighbouring transcripts. Subsequent gene set enrichment analysis will be used to examine changed pathways in disease development.
Third, to investigate aberrant tumor suppressor p63 promoter methylation to understand its role in the origin and development of Barrett's esophagus.

4.2 Methods

4.2.1 Establishing comparison groups

Comparison groups were established to interrogate methylation and associated expression data more comprehensively. The following groups were established:

- Normal (N) versus Barrett's esophagus without dysplasia (BE). No restrictions on dysplastic disease or EAC. For identification of Barrett's esophagus.
- Normal (N) versus high-grade dysplasia (HGD) AND esophageal adenocarcinoma (EAC). No restrictions on non-dysplastic BE or low-

grade dysplasia. For identification of intervention-requiring disease. Due to lack of restriction surrounding non-dysplastic BE; this is not necessarily a clinically valuable comparison: the majority of non-dysplastic BE patients will not progress to dysplastic disease or EAC[39] and as such, identification will result in costly surveillance programs without improvement in EAC prognosis.

- Normal (N) AND Barrett's esophagus without dysplasia (BE) versus high-grade dysplasia (HGD) AND esophageal adenocarcinoma (EAC). No restrictions on LGD. For differentiation of intervention-requiring disease from non-intervention requiring disease; this is the most clinically relevant comparison.

As a result of identification and classification challenges surrounding low grade dysplasia as well as much lower levels of homogeneity within sample biopsies; primary training analysis did not include LGD data. However, whole genome methylation and expression profiling was performed on these samples and weight given to shortlisted training regions hypermethylated in LGD. LGD samples were included in cohorts for targeted sequencing validation (Chapter 5: Target Region Validation and Characterization).

4.2.2 Genome-wide profiling

Genome-wide methylation and transcriptomic profiling was performed on DNA and RNA extracted from single tissue biopsies (n=48) from 27 patients comprising the Training cohort (Table 3-2).

4.2.2.1 Genome-wide methylation profiling

Clinical samples were profiled for methylation status using Illumina's Infinium HumanMethylation450 BeadChip® (HM450) in two rounds of 24 samples, performed as a service by the Australian Genome Research Facility (AGRF, Melbourne, Australia). DNA samples were re-checked by Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., Waltham, USA) and quality confirmed by resolution check on a 0.8% agarose gel (130V for 90min). 1µg of DNA was bisulfite converted using Zymo EZ DNA Methylation™ Kit as per

the manufacturer's protocol. Preliminary data analysis was performed using Illumina's GenomeStudio v2011.1 with Methylation module 1.9.0 software, using the default Illumina settings. Multiple rounds of array data were analysed (i) separately as discovery and validation sets, (ii) as a combined data set, strengthening the statistical significance of the assembled list of target regions.

4.2.2.2 Genome-wide transcriptome profiling

Clinical samples were profiled for comprehensive exploration of the transcriptome using Affymetrix GeneChip Human Transcriptome Array 2.0 (HTA2.0), in two rounds of 24 samples, performed as a service by the Ramaciotti Centre for Genomics (UNSW, Sydney, Australia). RNA samples were quality checked at three stages of processing (first cycle cRNA after clean up, second-cycle single-stranded cDNA after clean up and fragmented, labeled cDNA), checking yield, length and absorbance ratio 260nm/280nm at the first two stages and length at the third stage. 100ng total RNA input was used, with labeling performed using the Affymetrix GeneChip® WT PLUS Reagent Kit (Cat #902281) as per manufacturer's protocol. 5.2µg of fragmented and labeled ssDNA was hybridized to the array. Follow troubleshooting for poor hybridization, an additional pre-hybridization step was performed (30min), prior to incubation (16hrs at 45°C, rotation @ 60rpm, Incubator: Affymetrix 640). Fluidics as per protocol FS450_0001, Instrument: Affymetrix 450, Scanner: Affymetrix 3000 7G. Multiple rounds of array data were analysed (i) separately as discovery and validation sets, (ii) as a combined data set, strengthening the statistical significance of the assembled list of differentially expressed regions. Comparison groups established for whole-genome methylation profiling (Section 4.2.1 Establishing comparison groups) were maintained for comprehensive transcriptomic interrogation.

4.2.3 Statistical analysis

Bioinformatic analysis of genome-wide methylation and transcriptomic profiling was kindly performed by Dr Elena Zotenko, a computational analyst

with Professor Susan Clark's Epigenetics Research Program at the Garvan Institute of Medical Research.

4.2.3.1 Differential methylation statistical analyses

Illumina HumanMethylation450 raw data was pre-processed and background normalised with Bioconductor minfi package v3.2 (PMID: 24478339), using the preprocessIllumina(…, bg.correct = TRUE, normalise = "controls", reference=1) command. Resultant M-values were used for differential methylation statistical analysis[105]. Differential methylation analysis between comparison groups was carried out using the Bioconductor Linear Models for Microarray Data (LIMMA) package. Hyper-methylated probes were filtered based on false discovery rate (FDR < 0.05) and baseline methylation (avg. beta < 0.10), then merged into regions based on distance cut-off of 300bp. DMRs must span at least two filtered probes and at least one filtered probe with increase in methylation of at least 20%. Data was filtered to remove probes located on X and Y chromosomes. Regional filtering ensured identified loci are robust.

4.2.3.2 Differential expression statistical analyses

Affymetrix HTA2.0 raw data was pre-processed and robust multi-array average (RMA) normalised using Affymetrix Expression Console. Differential analysis between comparison groups was carried out using Affymetrix Transcriptome Analysis Console. Differentially expressed probes were selected based on FDR < 0.05 and log-fold change > +0.585 or -0.585 (equivalent to 1.5 fold change up or down). Data was filtered to remove probes located on X and Y chromosomes.

4.2.4 Integrative analysis: methylation and expression correlation

Genome-wide data were interrogated to investigate whether particular aberrant methylation was responsible for downstream altered RNA expression. Integrative analysis of HM450 and HTA2.0 array data was performed to identify regions whose hyper-methylation negatively correlated with expression of nearby transcripts (20kb cut-off from transcription start

site). Pearson correlation between expression and methylation beta-values of probes within hyper-methylated regions were computed. Region/transcript pairs showing strong negative correlation were given priority for validation by targeted amplicon sequencing.

In a separate script, the degree of correlation between expression and methylation in Normal, BE and HGD-EAC samples was measured. LGD samples were not included in this analysis due to sample heterogeneity and inter- as well as intra-observer disagreement in classification and composition. Towards this end, each transcript on the HTA2.0 array was paired with HM450 probes in the regions containing the transcript itself and +/- 20kb flanking regions. The degree of correlation was assessed with Spearman rank correlation test. For correlation analysis, we only considered pairs between highly variable transcripts and probes, which we define as 75th percentile of standard deviation in expression and methylation.

4.2.5 Gene Set Enrichment Analysis (GSEA)

Functional enrichment analysis was performed using Gene Set Enrichment Analysis (GSEA) against MolSigDB v5.0 database of known pathways, gene ontology terms and curated gene lists from the literature[106, 107]. The DAVID (Database for Annotation, Visualisation and Integrated Discovery) gene function annotation tool (http://david.ncifcrf.gov) was used to identify enriched biological themes and function-related gene groups occurring during disease development[108, 109].

4.2.6 Control samples

4.2.6.1 Control tissues

Proximal stomach and duodenal (small intestine, immediately beyond the stomach) tissues have similar mucosal structure to Barrett's esophageal tissue, enabling filtering of aberrant methylation indicative of tissue type (the maintenance of columnar epithelium phenotype) rather than the existence of BE. The training cohort for methylation and expression profiling included n=3 duodenal and n=3 proximal stomach controls; taken from non-dysplastic BE,

LGD and HGD patients. The external validation cohort (250 samples from 154 patients) included 21 gastric samples (from EAC patients) as well as 10 distal esophageal samples from patients suffering from gastroesophageal reflux disease (GERD) as controls. The question of biological similarity of matched normal proximal tissue from diseased patients (BE and EAC occur distally at the gastroesophageal junction) compared to that of healthy patients (proximal and distal esophageal tissue) was addressed by inclusion of a number of samples from these sites for methylation and expression profiling.

4.2.6.2 Control blood

To ensure clinically appropriate target region selection from methylation profiling, publicly available data (Gene Expression Omnibus (GEO) accession GSE48472)[110] were used to filter for methylation evident in normal peripheral blood (PT1-PT5). As described by Slieker et al[110], DNA was isolated from whole blood taken from five healthy volunteers (mean age 28 years, SD = 6.1) using the Qiagen mini kit according to manufacturer's protocol.

It must be noted that a more appropriate control blood sample would be the cell-free circulating DNA (cfc-DNA) fraction isolated from plasma, as this fraction is known to contain both mutations and DNA methylation patterns characteristic of disease occurring elsewhere in the body[11, 71, 72]. However due to the very low yields and highly fragmented DNA comprising the cell-free circulating fraction of plasma, it is not a practical control for genome-wide methylation profiling using HM450 technology. These considerations and the transition from tissue-identified putative biomarkers to blood plasma biomarkers are discussed more fully in Section 1.3.4 Blood-based biomarkers.

4.2.7 Aberrant promoter methylation of tumor suppressor p63

Loss of tumor suppressor p63 expression has been associated with carcinogenesis in a variety of cancer types[102]. This, and the observation that p63 null mouse models quickly develop Barrett's-like tissue[104] led us to interrogate p63 promoter methylation in our samples. We examined methylation in the two promoter regions of tumor suppressor p63 in the external validation cohort (250 samples from 154 patients (Table 3-2): 125 EAC, 18 BE, 85 non-tumor squamous esophagus and 21 normal stomach samples), plotting heat maps and bar charts for average methylation at HM450 probes within these regions.

4.3 Results

4.3.1 Quality control

4.3.1.1 Methylation profiling quality control

Illumina's Infinium HumanMethylation450 BeadChip® (HM450) was used for investigation of DNA methylation status, interrogating over 485 000 (485 512) CpG sites in the human methylome, representing ~1.7% total genomic CpG sites; a 17-fold improvement over the previous HumanMethylation27 BeadChip®[111, 112]. Furthermore, HM450 array data have shown to have excellent concordance with those from bisulfite sequencing[113], reinforcing it as one of the leading options for methylation analyses. Overall, forty-six (of 48) samples resulted in more than 485 000 detected CpG ($p<0.01$); at $p<0.05$ these two samples (484,993 and 462,934 at $p<0.01$) detected 485,081 and 475,120 CpG sites respectively. All internal sample-independent (staining, extension, target removal and hybridization) and sample-dependent (bisulfite conversion, specificity, non-polymorphic and negative) controls were as expected.

4.3.1.2 Transcriptomic profiling quality control

Recent advances in whole-transcript expression arrays such as Affymetrix's GeneChip® Human Transcriptome Array 2.0 (HTA 2.0) have significantly improved coverage of all transcript isoforms, providing significantly more

coverage than previous exon arrays, equivalent to two full lanes of RNA sequencing on an Illumina® HiSeq™ 2000 system. Affymetrix GeneChip® Human Transcriptome Array 2.0 (HTA 2.0) was used for gene expression profiling of all transcript isoforms, covering >285 000 full length transcripts (both coding and non-coding) to 44,699 protein coding and 22,829 non-protein coding genes. All samples were within the defined criteria for quality control at three stages of processing: cRNA, single-stranded cDNA and fragmented, labeled cDNA.

4.3.2 Global methylation density

Diagnostic density plots of beta values were assembled to check overall HM450 data quality and visualize global disease-associated methylation. Profiles of normal squamous tissue were examined to determine if distance aboral or esophageal disease present distally affects global methylation. Diagnostic density plots of samples from the metaplasia-dysplasia-adenocarcinoma sequence were examined globally to check for the presence of disease-associated aberrant methylation.

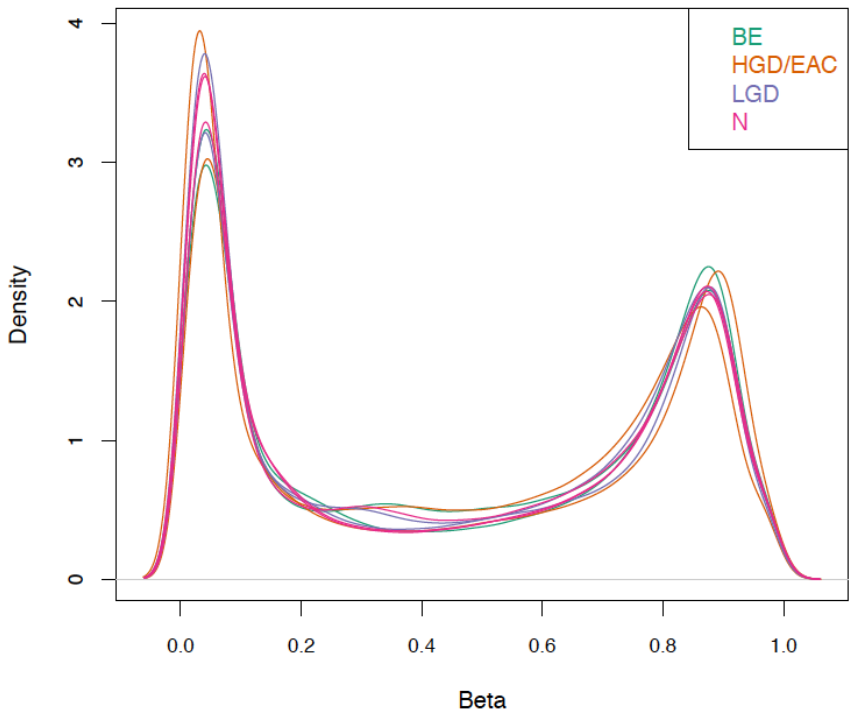
4.3.2.1 Normal squamous tissue methylation

Methylation density of normal squamous tissue biopsied proximally in the esophagus, typically 25cm aboral, was compared in samples taken from all disease classes (disease develops distally). There was no significant difference in the methylation profile of this tissue, irrespective of whether it was taken from a normal healthy patient, or a patient at any stage in the metaplasia-dysplasia-adenocarcinoma development sequence. This profile (Figure 4-1) is typical of normal, disease-free genome-wide methylation: the majority of regions are either entirely methylated or entirely unmethylated. Disease-associated aberrant methylation is associated with increased partial methylation, as can be seen in Figure 4-3.

There was also no significant difference in methylation profile of squamous esophageal tissue taken proximally or distally in a normal healthy patient. Comparison of global methylation beta values attained for proximal and distal

biopsies from the same (healthy, normal) individual show strong correlation (example Figure 4-2, Pearson correlation: 0.99). Thus, matched normal samples (taken proximally from diseased patients) can be used for further analyses.

(A)



(B)

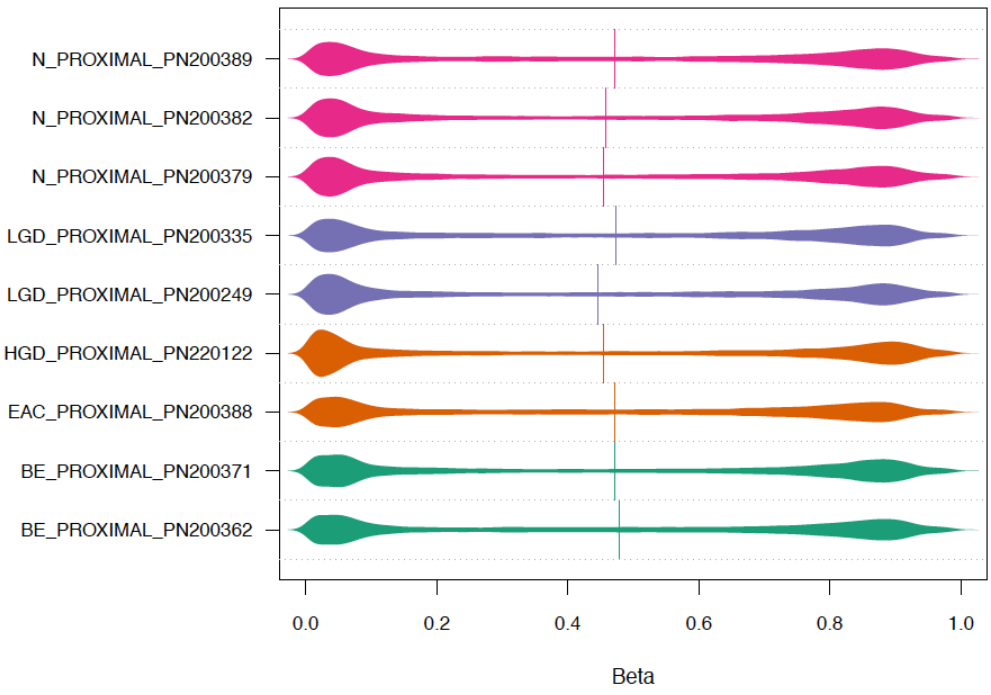


Figure 4-1: Comparison of methylation beta density in proximal esophageal tissue from examples of normal healthy patients versus patients at all stages of the metaplasia-dysplasia-adenocarcinoma sequence. Data represented as (A) Density and (B) Violin plots. All samples taken at 25cm proximally. Patient disease status as per Violin plot, y-axis. No significant difference was detected in methylation of proximal normal squamous epithelium, irrespective of patient disease status.

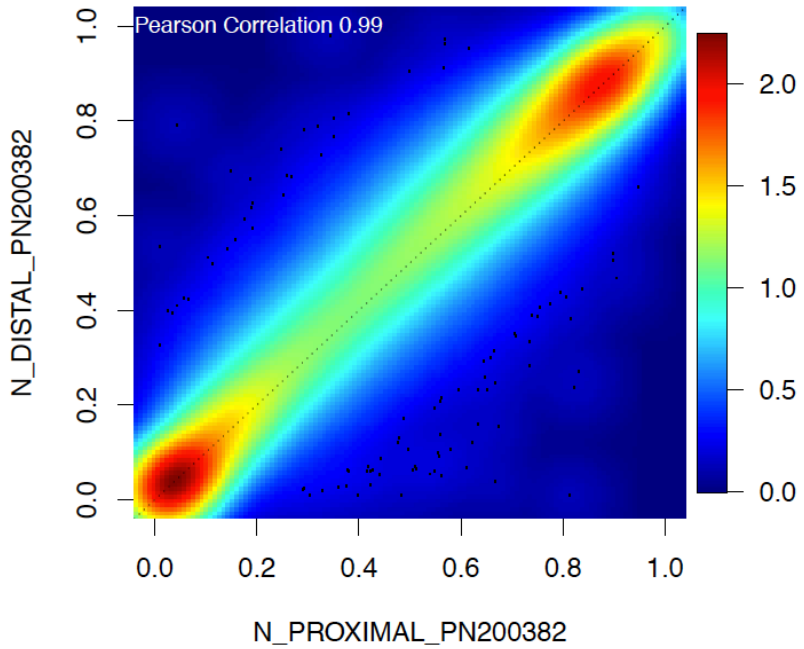


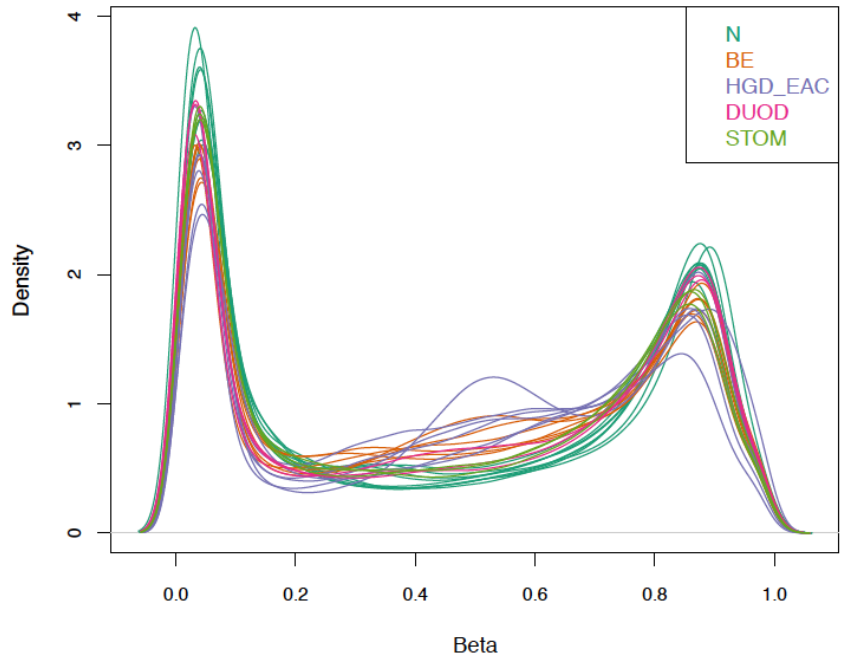
Figure 4-2: Correlation of global methylation beta values in normal squamous esophageal epithelium, taken proximally and distally from a single healthy patient. Strong correlation (Pearson correlation 0.99), supports hypotheses of uniform squamous epithelium methylation profile throughout the esophagus, irrespective of distance aboral.

4.3.2.2 Disease-associated aberrant methylation

Methylation density of distal, disease-associated epithelium was compared across all disease stages in the development of esophageal adenocarcinoma. As disease develops from normal squamous epithelium (healthy and diseased patient, biopsied proximally, Figure 4-3, in dark green) through the metaplasia-dysplasia-adenocarcinoma sequence, increased partial methylation is observed, indicating an increased prevalence of sub-

populations of cells becoming more highly methylated as disease progresses. As expected, methylation density profile of duodenal and proximal stomach tissues are similar to that of normal squamous epithelia from the esophagus: CpG sites are either universally methylated or entirely unmethylated across the entire cell population.

(A)



(B)

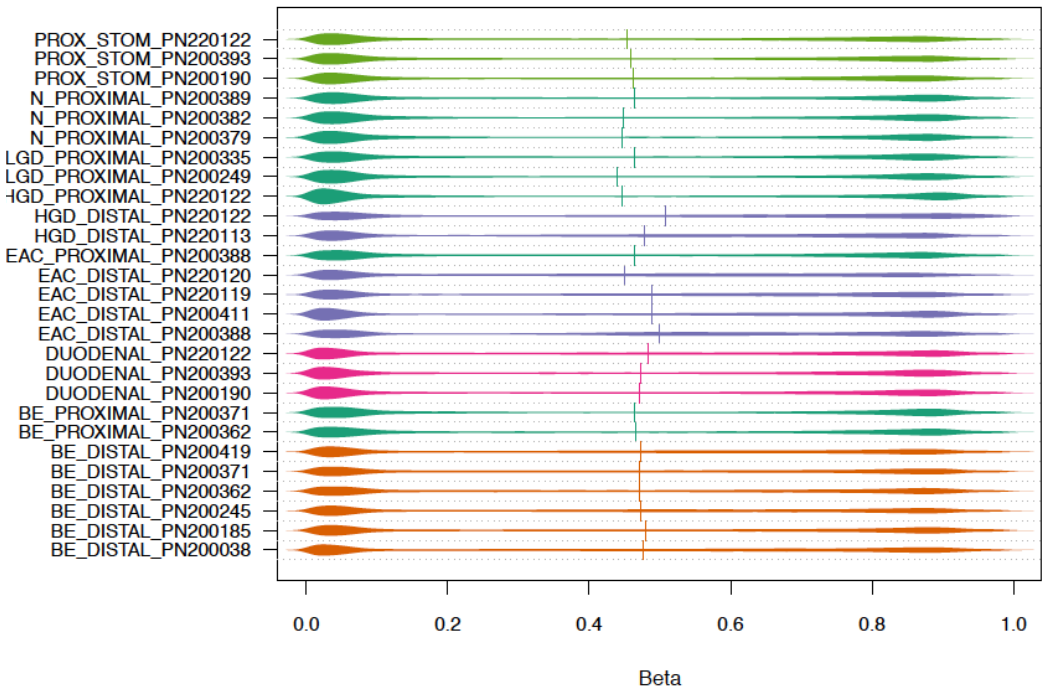
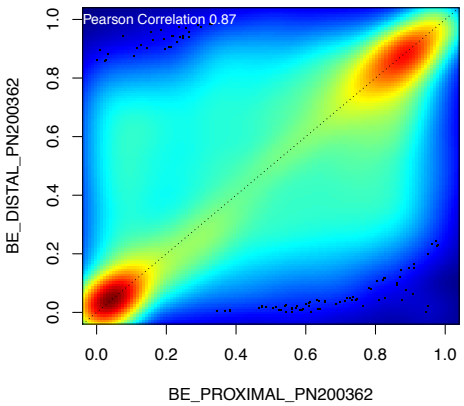
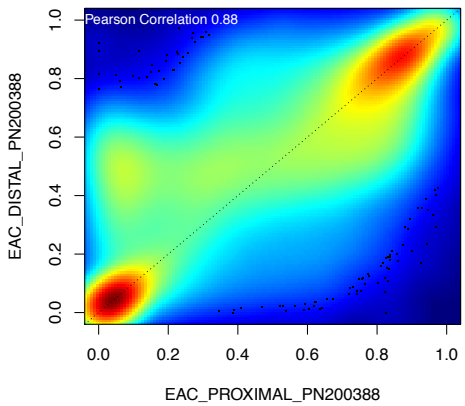


Figure 4-3: Comparison of methylation beta density in distal esophageal tissue from normal healthy patients versus patients at all stages of the metaplasia-dysplasia-adenocarcinoma sequence. Data represented as (A) Density and (B) Violin plots. Increased densities of intermediate beta-values are observed for diseased samples with respect to normal squamous esophageal, duodenal and proximal stomach epithelium; showing positive correlation with disease severity.

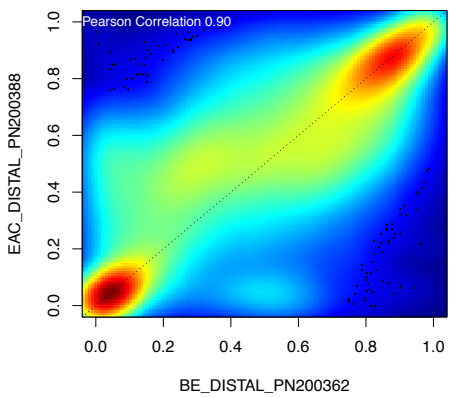
Disease-associated aberrant methylation is apparent between all disease classes, including the three comparison classes used for this study: N v BE, N v HGD-EAC and BE v HGD-EAC (Figure 4-4 (i), (ii), (iii)). Of clinical interest is detectable differential methylation between non-dysplastic BE and BE with low-grade dysplasia (Figure 4-4 (iv), Pearson correlation 0.95), a problematic distinction by histology.



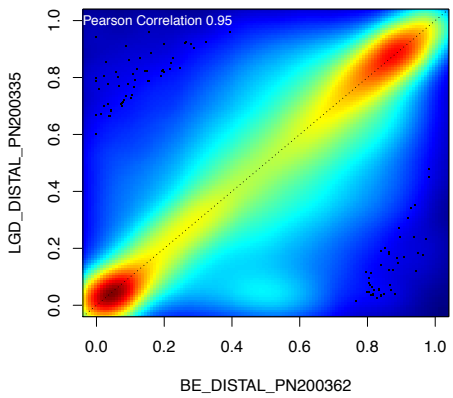
(i)



(ii)



(iii)



(iv)

Figure 4-4: Correlation of global methylation beta values between disease classes (i) N v BE, N is matched normal sample taken proximally. Pearson correlation 0.87; (ii) N v HGD-EAC, N is matched normal sample taken proximally. Pearson correlation 0.88; (iii) BE v HGD-EAC, samples from different patients, Pearson correlation 0.90; (iv) Non-dysplastic BE v BE-LGD, samples from different patients, Pearson correlation 0.95. These plots show evidence of differential methylation across all disease comparisons, including the clinically significant development of LGD from non-dysplastic disease.

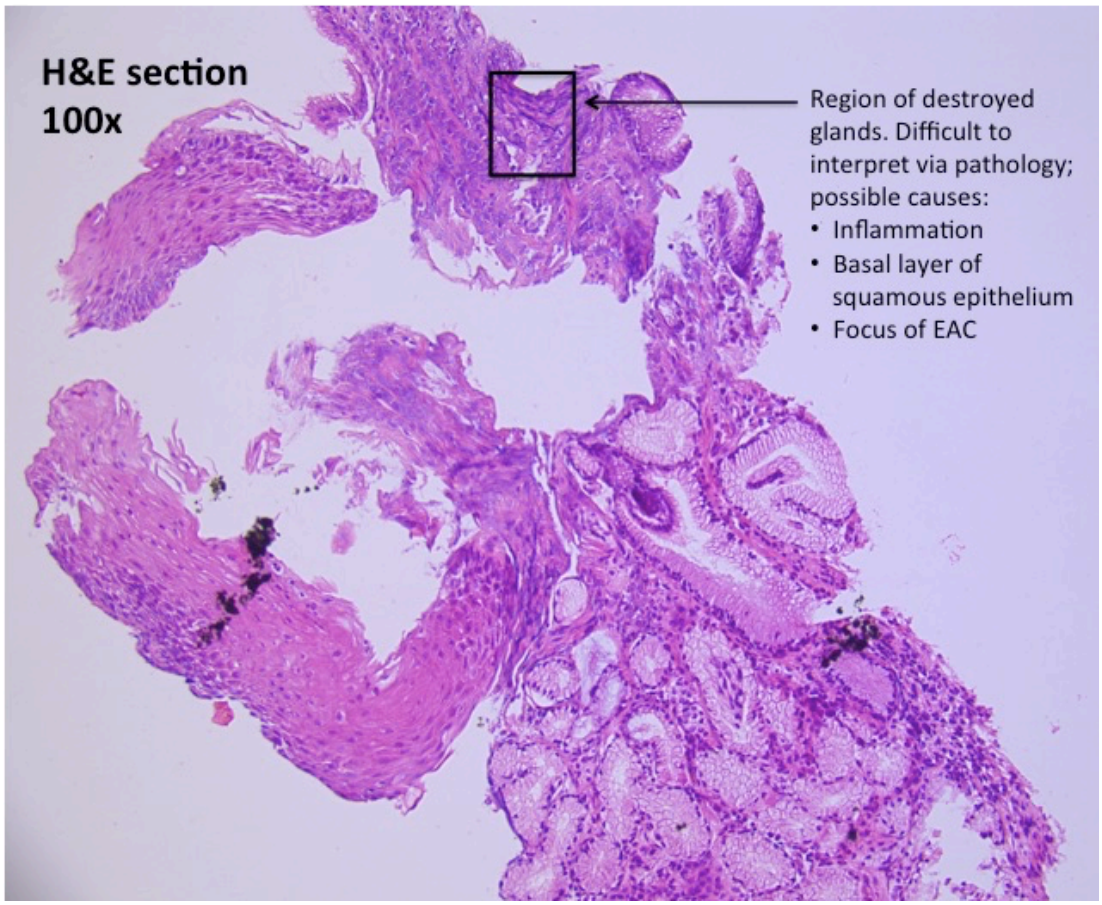
4.3.3 Tissue heterogeneity

Tissue heterogeneity is problematic for current diagnostic methods, which require not only accurate disease identification and biopsy by the endoscopist, but rely on representative sectioning (blinded, it is not possible to identify disease visually in the biopsy) for histology to identify heterogeneous disease. One interesting sample came from a 70-year old male, taken distally at 40cm (at the GEJ). The biopsy contained a small area of destroyed glands, possibly a focus of adenocarcinoma, evaluated as such by 1 of 4 pathologists in independent, blinded evaluation (Figure 4-5). Interestingly, on repeat analysis, this same pathologist identified that dysplasia was present, but instead recorded a much higher composition of low-grade dysplasia. The difficulty faced by pathologists with a sample such as this is, despite universal recognition of an area of destroyed glands, there is insufficient information to determine whether the area of destroyed glands is due to inflammation, adenocarcinoma, or is simply the basal layer of squamous epithelium.

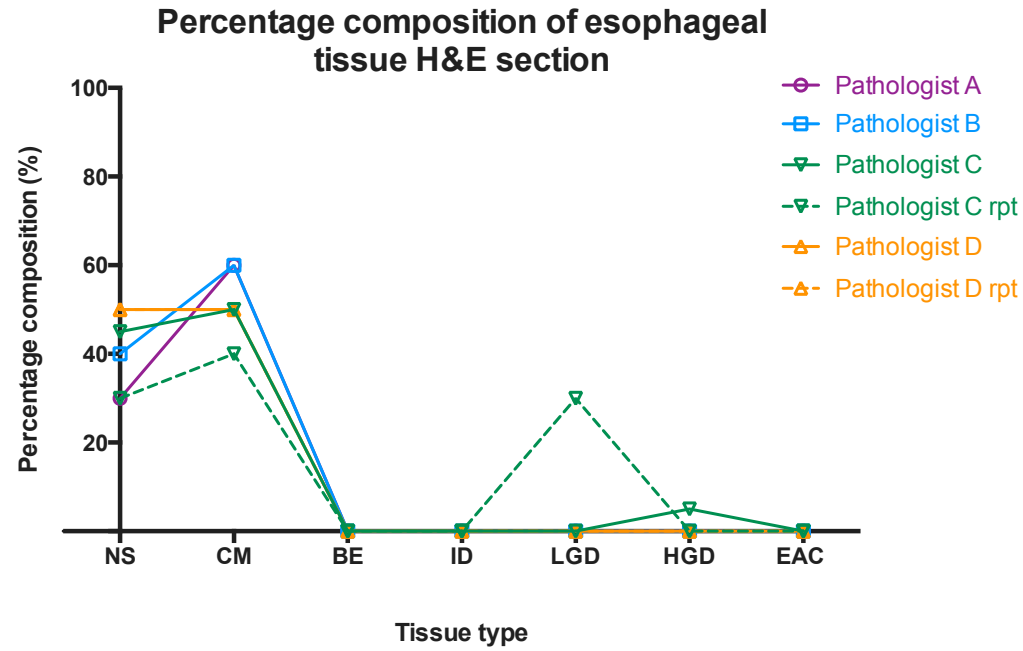
Assuming the H&E section to be representative of the entire biopsy, this very small section of destroyed glands, despite being swamped by an abundance of copies of DNA from surrounding normal (squamous and cardiac mucosa) epithelium, was able to be identified as focal adenocarcinoma by methylation changes apparent when comparing beta values from this sample with that of a matched normal sample taken proximally at the same visit. A proportion of the observed aberrant methylation (Figure 4-5(C)) is hypothesized to be

attributable to a small focus of adenocarcinoma; supported by slightly hypermethylated probes detected for this sample in regions identified as aberrantly methylated in the group of HGD-EAC samples profiled. Figure 4-5(D) shows five regions, hypermethylated in HGD and EAC samples (LRRC43, CACNA2D2, ZNF221, ISM2 and TRANK1) also showing increased methylation in this sample. It is also plausible that a proportion of aberrant methylation may also be due to tissue type differences between proximal squamous epithelium and cardiac mucosa at the GEJ, however the presence of methylation in regions hypermethylated in dysplastic and adenocarcinoma samples supports the hypothesis that much of the aberrant methylation in this case is attributable to the presence of focal disease, rather than maintenance of columnar-type mucosa.

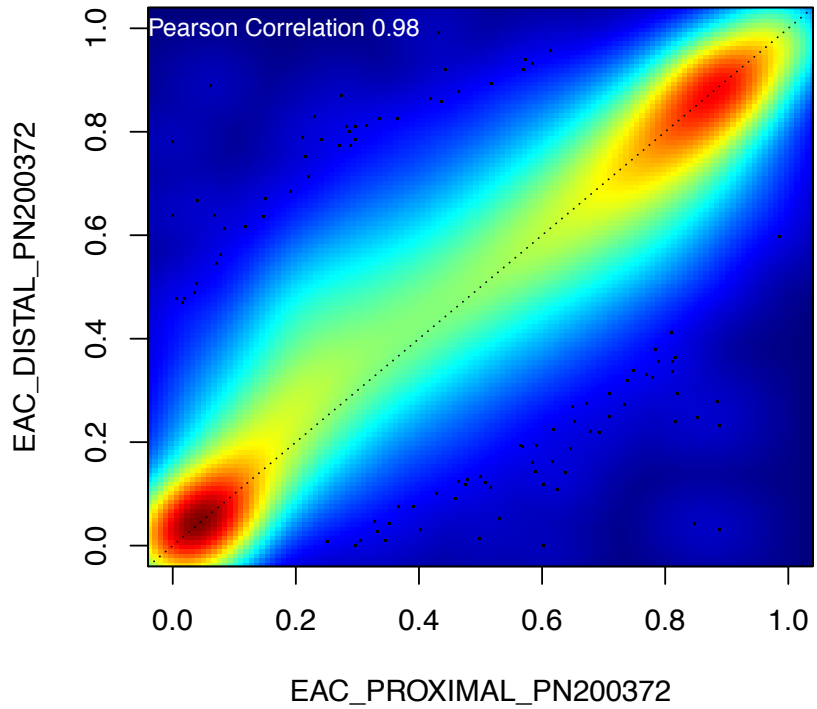
(A)



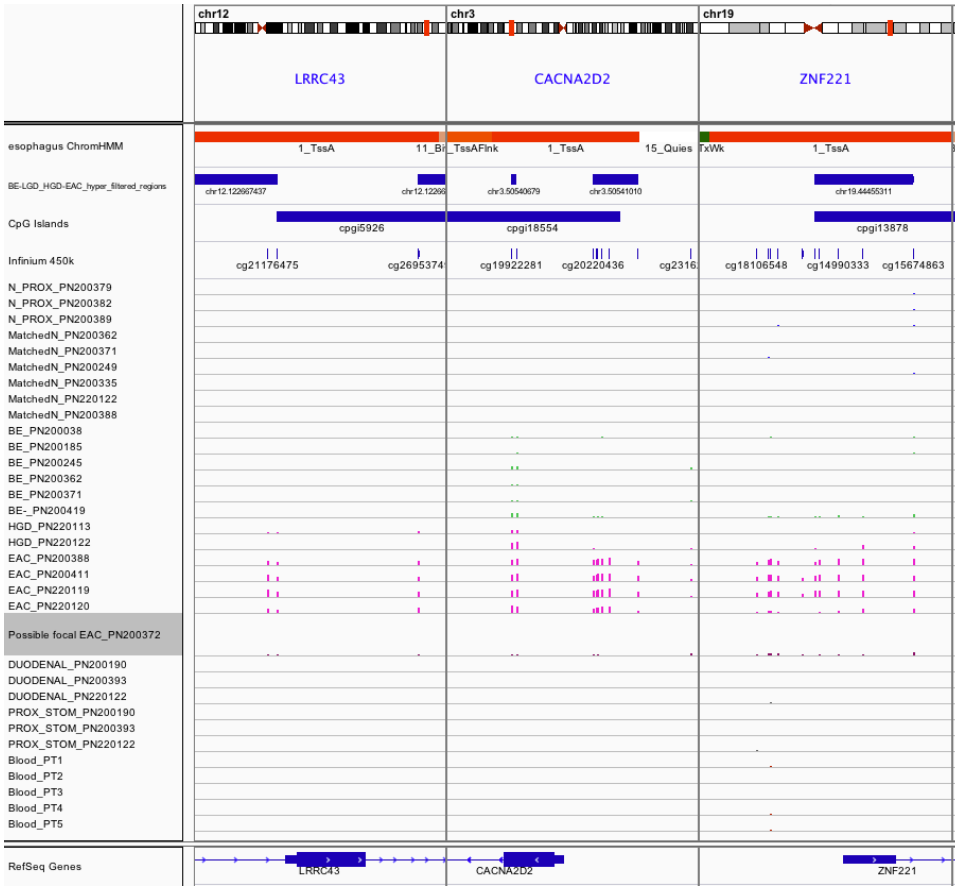
(B)


Percentage composition of esophageal tissue H&E section
Pathologist A
Pathologist B
Pathologist C
Pathologist C rpt
Pathologist D
Pathologist D rpt
Percentage composition (%)
100
80
60
40
20
NS
CM
BE
ID
LGD
HGD
EAC
Tissue type

(C)


Pearson Correlation 0.98
EAC_DISTAL_PN200372
0.0
0.2
0.4
0.6
0.8
1.0
EAC_PROXIMAL_PN200372

(D)


chr12
chr3
chr19
LRRC43
CACNA2D2
ZNF221
esophagus ChromHMM
1_TssA
15_Quies
BE-LGD_HGD-EAC_hyper_filtered_regions
CpG Islands
cpgi5926
cpgi18554
cpgi13878
Infinium 450k
cg21176475
cg2695374
cg19922281
cg20220436
cg18106548
cg14990333
cg15674863
N_PROX_PN200379
N_PROX_PN200382
N_PROX_PN200389
MatchedN_PN200362
MatchedN_PN200371
MatchedN_PN200249
MatchedN_PN200335
MatchedN_PN220122
MatchedN_PN200388
BE_PN200038
BE_PN200185
BE_PN200245
BE_PN200362
BE_PN200371
BE-_PN200419
HGD_PN220113
HGD_PN220122
EAC_PN200388
EAC_PN200411
EAC_PN220119
EAC_PN220120
Possible focal EAC_PN200372
DUODENAL_PN200190
DUODENAL_PN200393
DUODENAL_PN220122
PROX_STOM_PN200190
PROX_STOM_PN200393
PROX_STOM_PN220122
Blood_PT1
Blood_PT2
Blood_PT3
Blood_PT4
Blood_PT5
RefSeq Genes
LRRC43
CACNA2D2
ZNF221

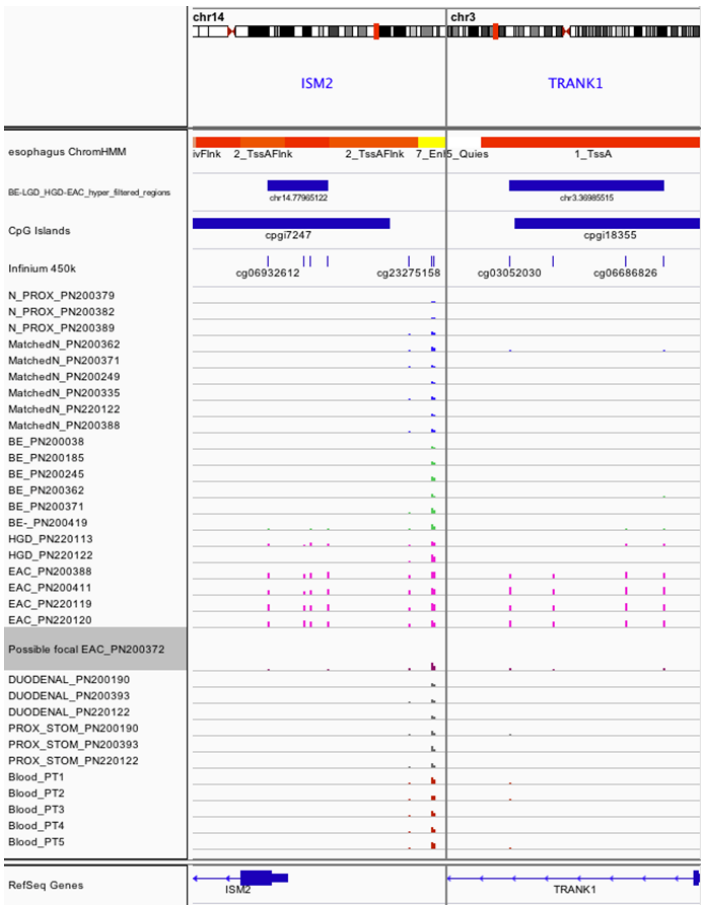

chr14
chr3
ISM2
TRANK1
esophagus ChromHMM
1_TssA
BE-LGD_HGD-EAC_hyper_filtered_regions
CpG Islands
cpgi7247
cpgi18355
Infinium 450k
cg06932612
cg23275158
cg03052030
cg06686826
N_PROX_PN200379
N_PROX_PN200382
N_PROX_PN200389
MatchedN_PN200362
MatchedN_PN200371
MatchedN_PN200249
MatchedN_PN200335
MatchedN_PN220122
MatchedN_PN200388
BE_PN200038
BE_PN200185
BE_PN200245
BE_PN200362
BE_PN200371
BE-_PN200419
HGD_PN220113
HGD_PN220122
EAC_PN200388
EAC_PN200411
EAC_PN220119
EAC_PN220120
Possible focal EAC_PN200372
DUODENAL_PN200190
DUODENAL_PN200393
DUODENAL_PN220122
PROX_STOM_PN200190
PROX_STOM_PN200393
PROX_STOM_PN220122
Blood_PT1
Blood_PT2
Blood_PT3
Blood_PT4
Blood_PT5
RefSeq Genes
ISM2
TRANK1

Figure 4-5: Focal adenocarcinoma, difficult to diagnose by histopathologic evaluation, can be detected by aberrant methylation. (A) H&E stained section (100x) of an esophageal tissue biopsy (40cm distally at GEJ) taken from a 70 year old male with moderately differentiated adenocarcinoma at the gastroesophageal junction and lymph node and liver metastases. Area of destroyed glands (boxed), identified histologically as a possible small focus of adenocarcinoma. (B) Percentage composition classification from four independent, blinded pathologists. Pathologist C and D unknowingly evaluated the same sample a second time, a minimum of 5 weeks later. All pathologists agreed the predominant tissue type was a mixture of cardiac mucosa and normal squamous epithelium. Only 1 of 4 pathologists identified the area of destroyed glands as focal dysplasia/adenocarcinoma. NS: normal squamous, CM: cardiac mucosa, BE: Barrett's esophagus, ID: indefinite for dysplasia, LGD: low-grade dysplasia, HGD: high-grade dysplasia, EAC: esophageal adenocarcinoma. (C) Correlation of global methylation beta values between the heterogeneous distal biopsy (possible focal adenocarcinoma, EAC_DISTAL_PN200372) and a matched normal tissue biopsy taken proximally at 25cm (EAC_PROXIMAL_PN200372). Pearson correlation 0.98 indicates methylation profile is very similar between the two samples, however aberrant hypermethylation in the distal sample (beta ~0.50) is discernable by skew. (D) Low level aberrant disease-associated hypermethylation of LRRC43, CACNA2D2, ZNF221, ISM2 and TRANK1 is detectable in this particular sample (Possible focal EAC_PN200372, in grey), despite the associated heterogeneity problems. Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

4.3.4 Global methylation and transcriptomic analyses

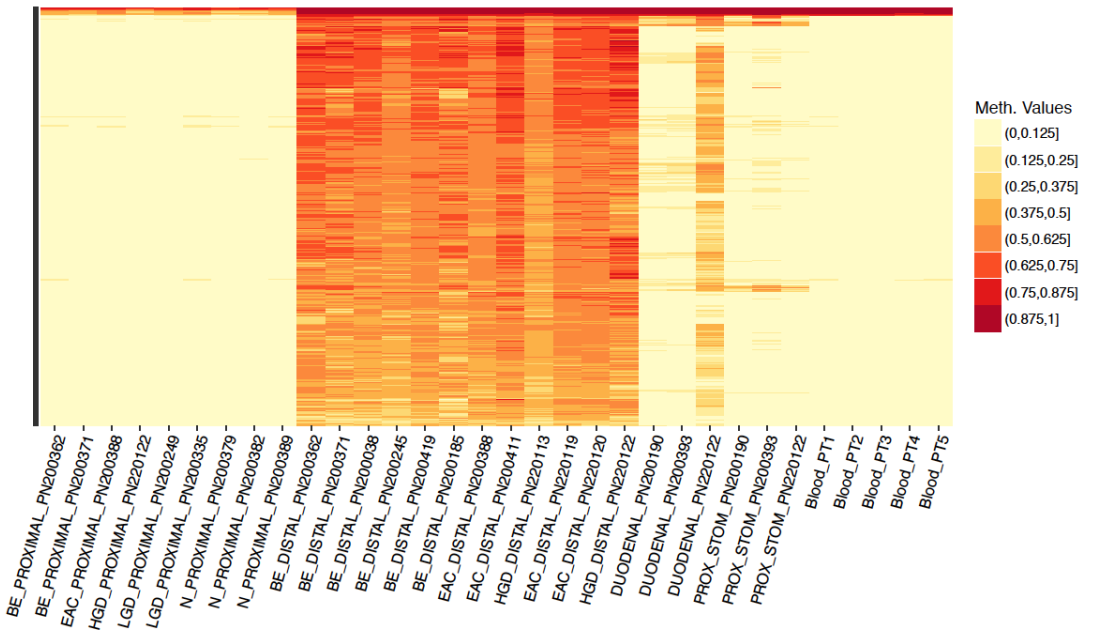
4.3.4.1 Genome-wide methylation profiling

As expected, the largest differential methylation was detected in HGD-EAC when compared to N (55,424 hypermethylated (Figure 4-6) and 20,876 hypomethylated (Figure 4-7) CpG sites; FDR < 0.01 and $\Delta\beta \geq 0.20$), followed by BE compared to N (44,201 hypermethylated (Figure 4-8) and 26,985

hypomethylated (Figure 4-9) CpG sites; FDR < 0.01 and Δβ ≥ 0.20) and finally BE compared to HGD-EAC (1,648 hypermethylated (Figure 4-10) and 153 hypomethylated (Figure 4-12) CpG sites; FDR< 0.01 and Δβ ≥ 0.20). Details of the top 100 hyper- and hypomethylated probes detected for each comparison group: N v BE, N v HGD-EAC and BE v HGD-EAC are given in Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling.

The most significantly hypermethylated regions in HGD and EAC (with respect to normal esophageal epithelium) were, in the most part, also significantly hypermethylated in non-dysplastic BE, but unmethylated in peripheral blood from normal, healthy patients, as well as normal esophageal, duodenal and proximal stomach mucosa (Figure 4-6(A)). Interestingly, one of three duodenal tissue samples did not cluster with other duodenal, proximal stomach and normal esophageal mucosa, instead showing slight hypermethylation in these regions. This sample is shown as an outlier in Figure 4-6(B).

(A) N v HGD-EAC Hypermethylation heat map



(B) N v HGD-EAC Hypermethylation PCA plot

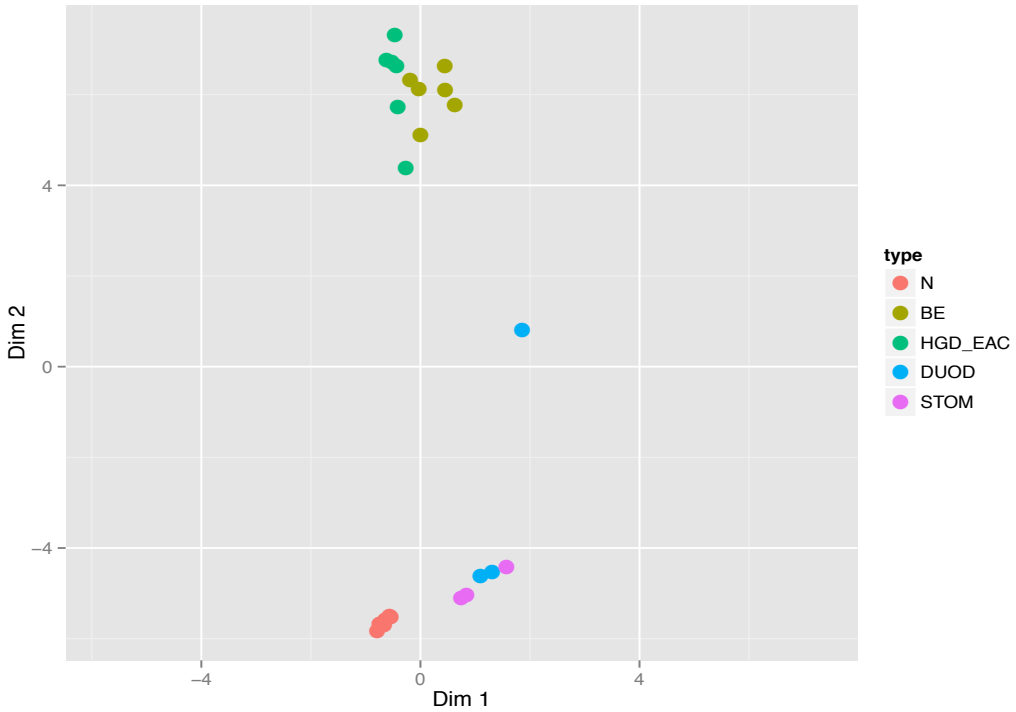
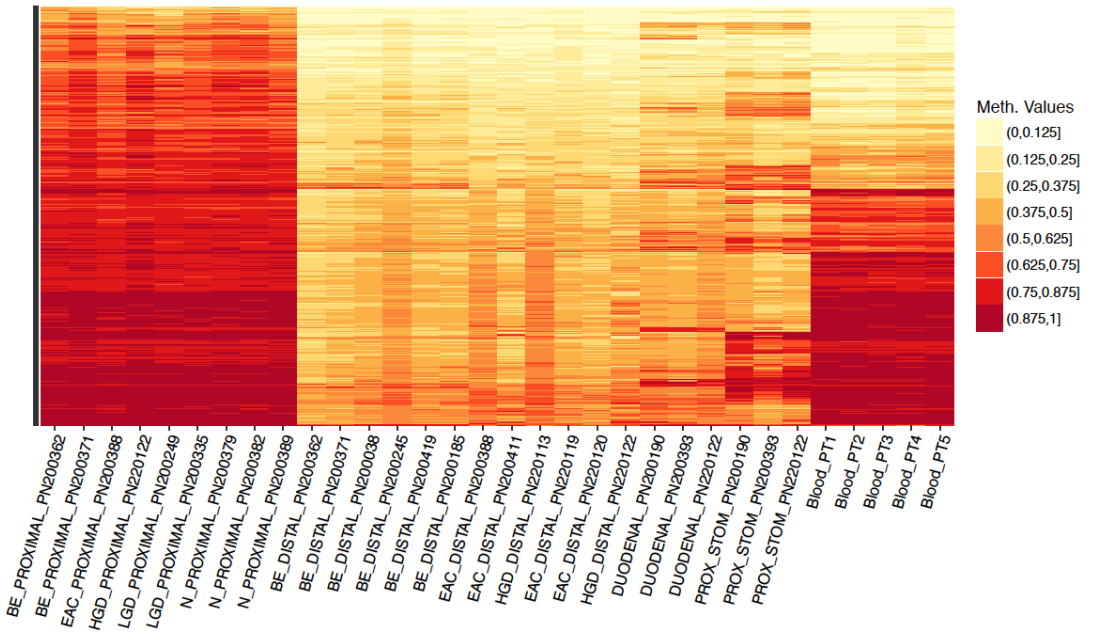


Figure 4-6: Top 500 hypermethylated CpG sites detected in the N v HGD-EAC comparison, FDR < 0.01 and Δβ ≥ 0.20 expressed as a (A) Heatmap showing patients (and their respective classification), with methylation values assigned as eight discrete regions from 0.000 – 1.000 in increments of 0.125 for better visualization (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

The most significantly hypomethylated regions in HGD and EAC (with respect to normal esophageal epithelium) were also hypomethylated in non-dysplastic BE. More than half of the top 500 identified regions were also hypomethylated with respect to normal peripheral blood (Section 4.2.6.2 Control blood), however a significant number of regions displayed little or no methylation in normal peripheral blood; and thus unable to be differentiated from HGD/EAC (Figure 4-7(A)). For this comparison, normal mucosa samples cluster strongly, showing very similar methylation profiles. Duodenal samples cluster with non-dysplastic, dysplastic and EAC samples; with more variably methylated proximal stomach samples clustering alone (Figure 4-7(B)). From this, we can hypothesize that much of the hypomethylation occurring is attributable to the maintenance of columnar mucosa phenotype

rather than disease-associated progression through the metaplasia-dysplasia-adenocarcinoma sequence. For this reason, no hypomethylation targets were carried through to validation and targeted sequencing studies.

(A) N v HGD-EAC Hypomethylation heat map



(B) N v HGD-EAC Hypomethylation PCA plot

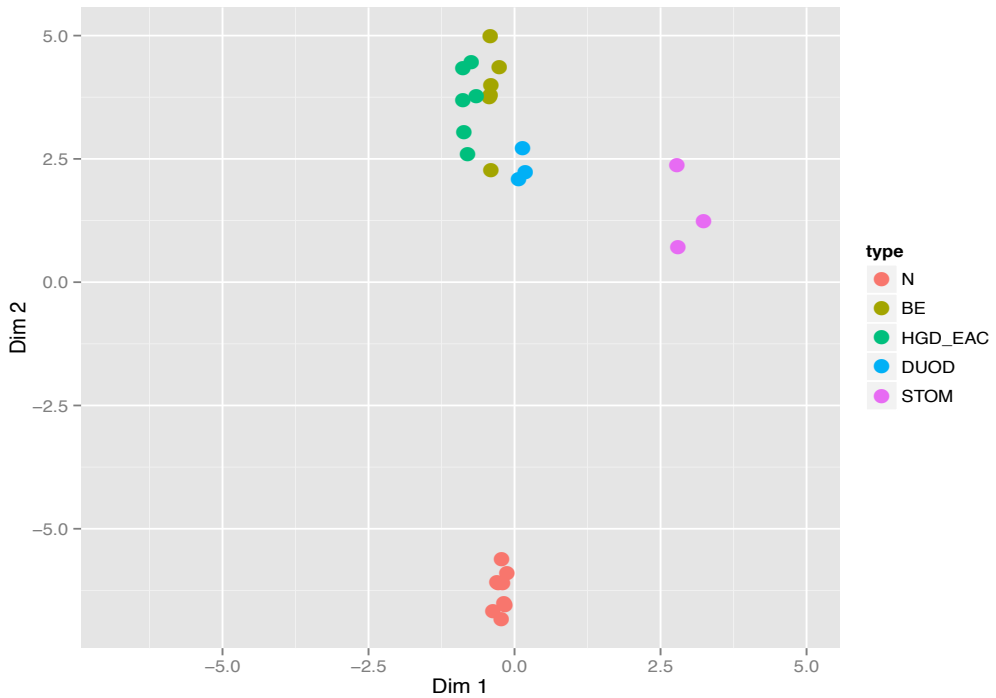
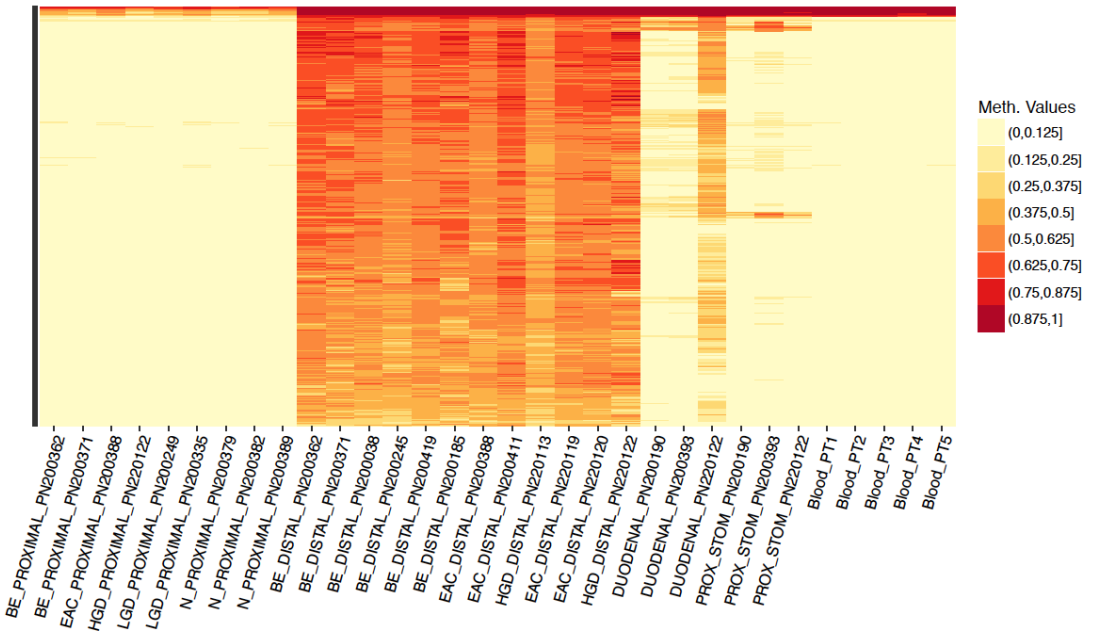


Figure 4-7: Top 500 hypomethylated CpG sites detected in the N v HGD-EAC comparison, FDR < 0.01 and Δβ ≥ 0.20 expressed as a (A) Heatmap

showing patients (and their respective classification), with methylation values assigned as eight discrete regions from 0.000 – 1.000 in increments of 0.125 for better visualization (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

Observations for hypermethylation in N v BE are very similar to those for hypermethylation in N v HGD-EAC. The most significantly hypermethylated regions in BE (with respect to normal esophageal epithelium) were, in the most part, also significantly hypermethylated in HGD and EAC, but unmethylated in peripheral blood from normal, healthy patients, as well as normal esophageal, duodenal and proximal stomach mucosa (Figure 4-8(A)). As for the N v HGD-EAC comparison, one of three duodenal tissue samples did not cluster with other duodenal, proximal stomach and normal esophageal mucosa, instead showing slight hypermethylation in these regions. This sample is shown as an outlier in Figure 4-8(B).

(A) N v BE Hypermethylation heat map



(B) N v BE Hypermethylation PCA plot

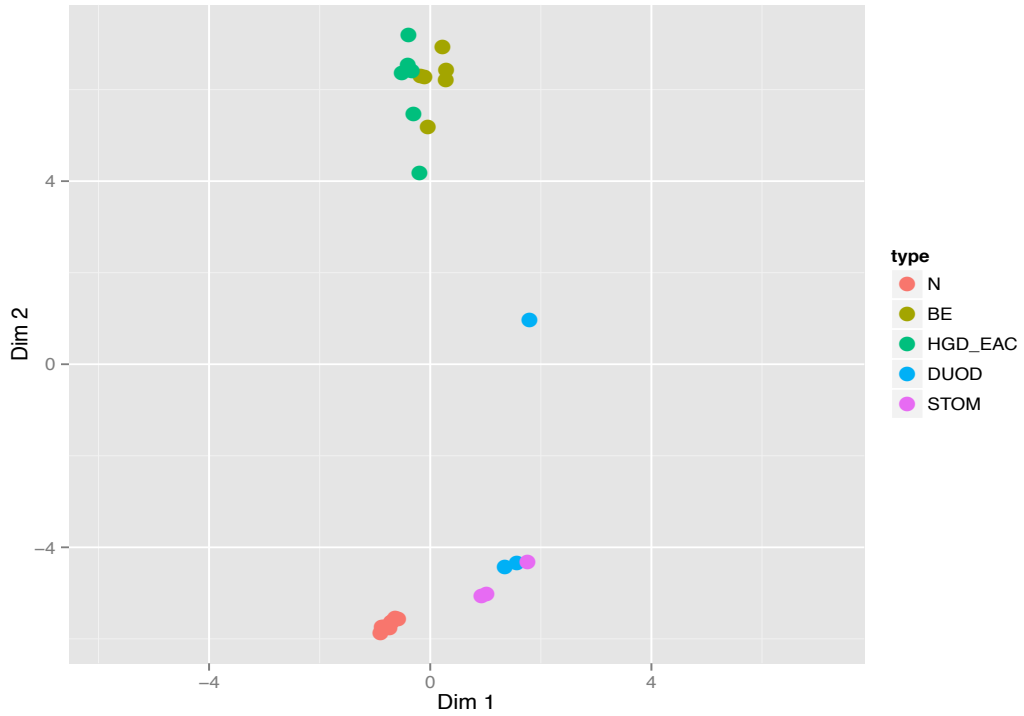
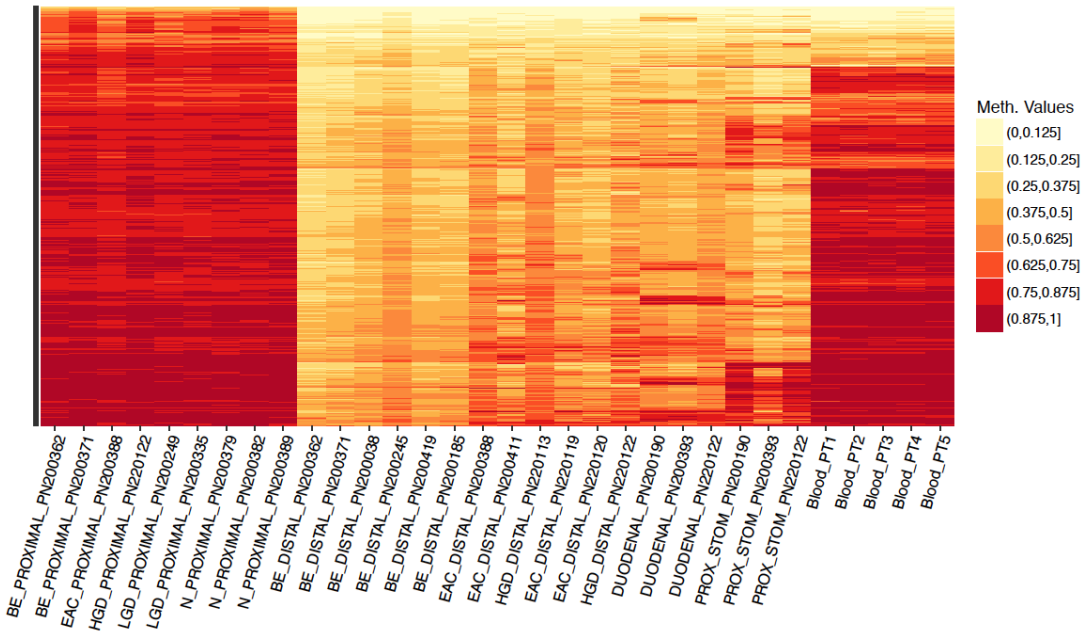


Figure 4-8: Top 500 hypermethylated CpG sites detected in the N v BE comparison, FDR < 0.01 and Δβ ≥ 0.20 expressed as a (A) Heatmap showing patients (and their respective classification), with methylation values assigned as eight discrete regions from 0.000 – 1.000 in increments of 0.125 for better visualization (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

Observations for hypomethylation in N v BE are very similar to those for hypomethylation in N v HGD-EAC, with the exception that slightly higher methylation levels were observed in duodenal, proximal stomach, dysplastic disease and EAC samples. Also, there were fewer incidences of regions with little or no methylation in normal peripheral blood (Figure 4-9(A)). Clustering for this comparison is very similar to that of hypomethylation in N v HGD-EAC with normal mucosa clustering strongly, indicating very similar inter-patient methylation profile. Duodenal samples cluster with non-dysplastic, dysplastic and EAC samples; with more variably methylated proximal stomach samples clustering alone (Figure 4-9(B)). These data further support the hypothesis of hypomethylation being attributable to maintenance of columnar mucosa rather than disease-associated changes. For this

reason, no hypomethylation targets were carried through to validation and targeted sequencing studies.

(A) N v BE Hypomethylation heat map



(B) N v BE Hypomethylation PCA plot

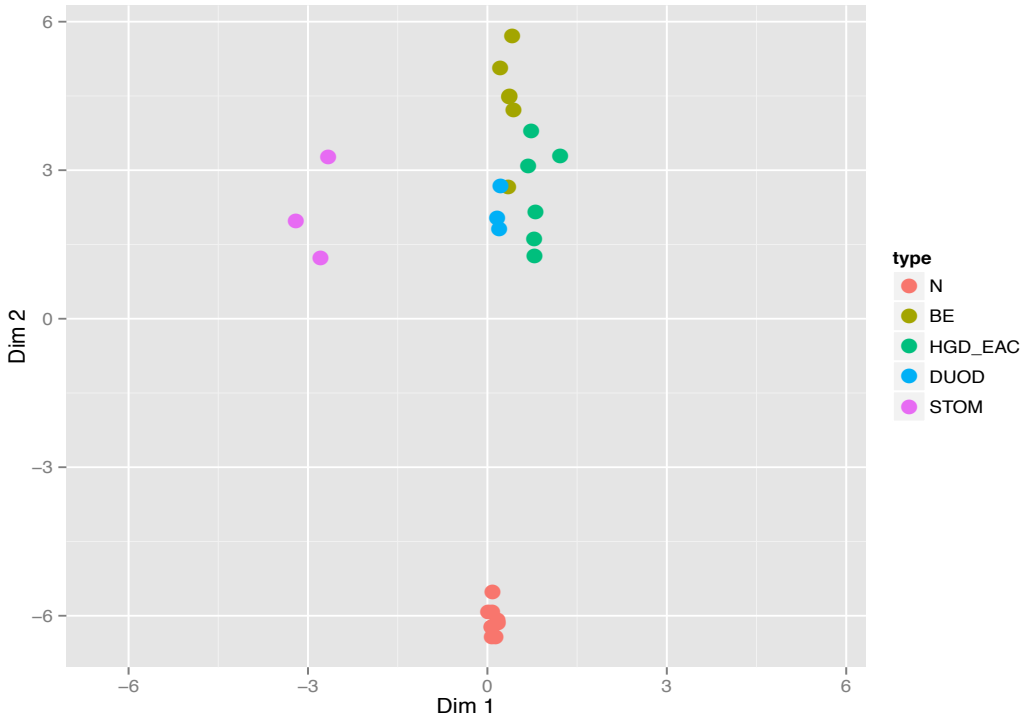


Figure 4-9: Top 500 hypomethylated CpG sites detected in the N v BE comparison, FDR < 0.01 and Δβ ≥ 0.20 expressed as a (A) Heatmap showing patients (and their respective classification), with methylation values

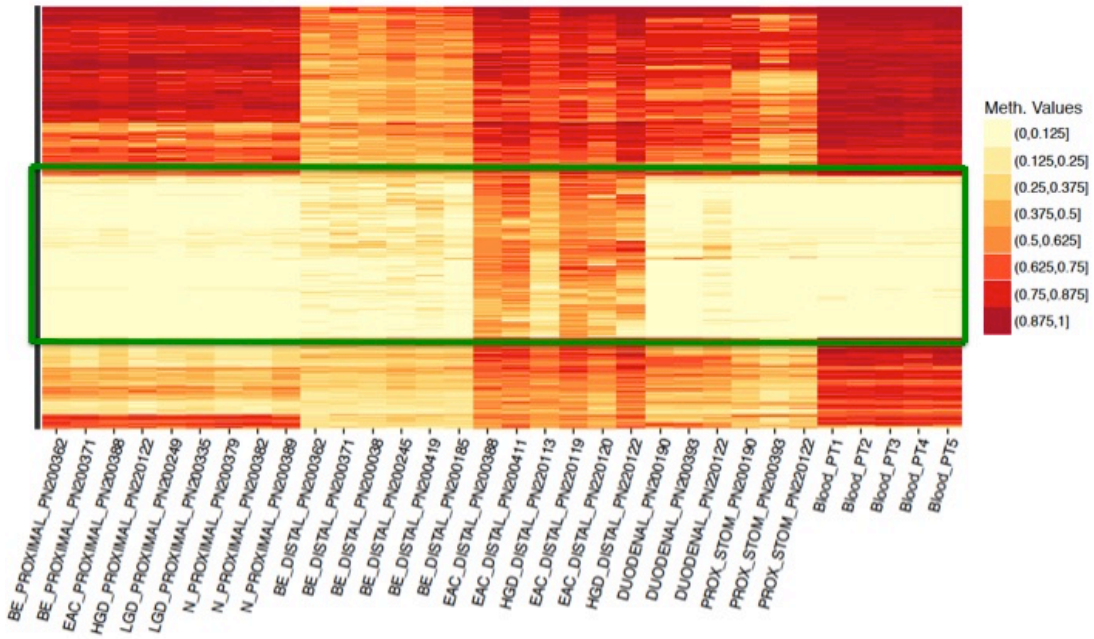
assigned as eight discrete regions from 0.000 – 1.000 in increments of 0.125 for better visualization (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

Results for hypermethylation in the BE v HGD-EAC comparison showed several variable subsets (Figure 4-10(A)). Of clinical interest is the subset of hypermethylated HGD-EAC samples with corresponding absence of methylation in non-dysplastic BE, normal esophageal mucosa, control tissues and blood (marked by a green box). It is these targets that will be the focus of further validation as they are able to differentiate intervention-requiring disease from the relatively ubiquitous, non-dysplastic, benign, Barrett's esophagus.

Below the green box (Figure 4-10(A)) is a similar subset with the exception of high-level methylation in normal peripheral blood. This highlights the importance of using normal peripheral blood as a control for biomarkers with potential transferability for a screening blood test: regions corresponding to these CpG sites are not clinically useful as HGD-EAC is not discernable from the methylation occurring in the blood of normal, healthy individuals.

Clustering for this comparison shows distinct clustering into tissue type and disease classes (Figure 4-10(B)).

(A) BE v HGD-EAC Hypermethylation heat map



(B) BE v HGD-EAC Hypermethylation PCA plot

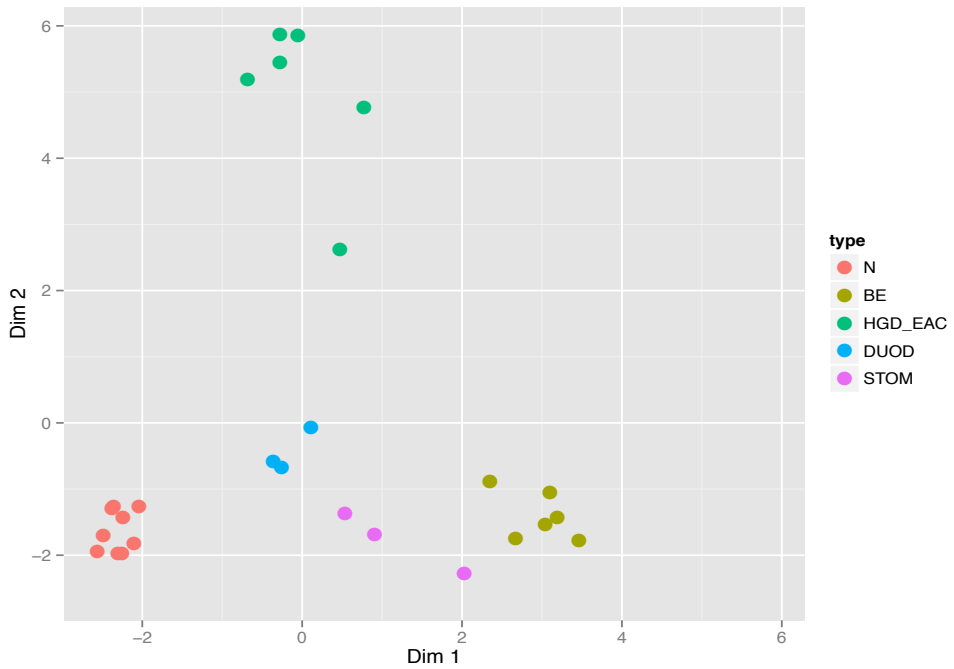


Figure 4-10: Top 500 hypermethylated CpG sites detected in the BE v HGD-EAC comparison, FDR < 0.01 and Δβ ≥ 0.20 expressed as a (A) Heatmap showing patients (and their respective classification), with methylation values assigned as eight discrete regions from 0.000 – 1.000 in increments of 0.125

for better visualization (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

The subset above the green box (Figure 4-10(A)) is interesting in that increased methylation is observed across all classifications (HGD-EAC, normal esophageal mucosa, control tissues and normal peripheral blood) with the exception of lower levels of methylation in non-dysplastic BE (note that this is low-level methylation, rather than absence of methylation). Regions from this subset were not considered for further validation due to inability to differentiate between normal and cancerous tissues, however they present an interesting case. What is happening biologically for de-methylation to occur only during metaplasia, returning strongly to a methylated status if progression to EAC occurs? It is interesting to note that these regions are also strongly methylated in blood. Many CpG sites from this subset correspond to regions not annotated by Gencode v19. However, those that do, correspond to protein coding genes such as FOXK1, CANT1, GPR37L1. Interestingly, these regions often are quite small, surrounded by regions of universal methylation (irrespective of disease status), and occurring within the gene body, for example in FOXK1 and CANT1. However there are instances, such as in GPR37L1 that this BE-specific reduced methylation is occurring in a promoter region.

An example of a protein coding gene, Forkhead box K1 (FOXK1), that contains a region of decreased methylation in only non-dysplastic Barrett's mucosa is given Figure 4-11 in below. Average methylation of normal esophageal mucosa for this region is 0.969; normal peripheral blood 0.939 and esophageal adenocarcinoma 0.862 (slightly lower due to a single HGD-EAC sample with reduced methylation), whereas BE tissues averaged 0.491. FOXK1 has been proposed to participate in processes such as cell differentiation, regulation of transcription and multicellular organismal development. FOXK1 gene ontology (GO) annotations include transcription factor activity, sequence-specific DNA binding and RNA polymerase II transcription factor activity. Therefore, we can hypothesize that these lower levels of methylation in metaplastic tissue may be playing a role in cell

differentiation processes specific to Barrett's esophagus mucosal formation and maintenance.

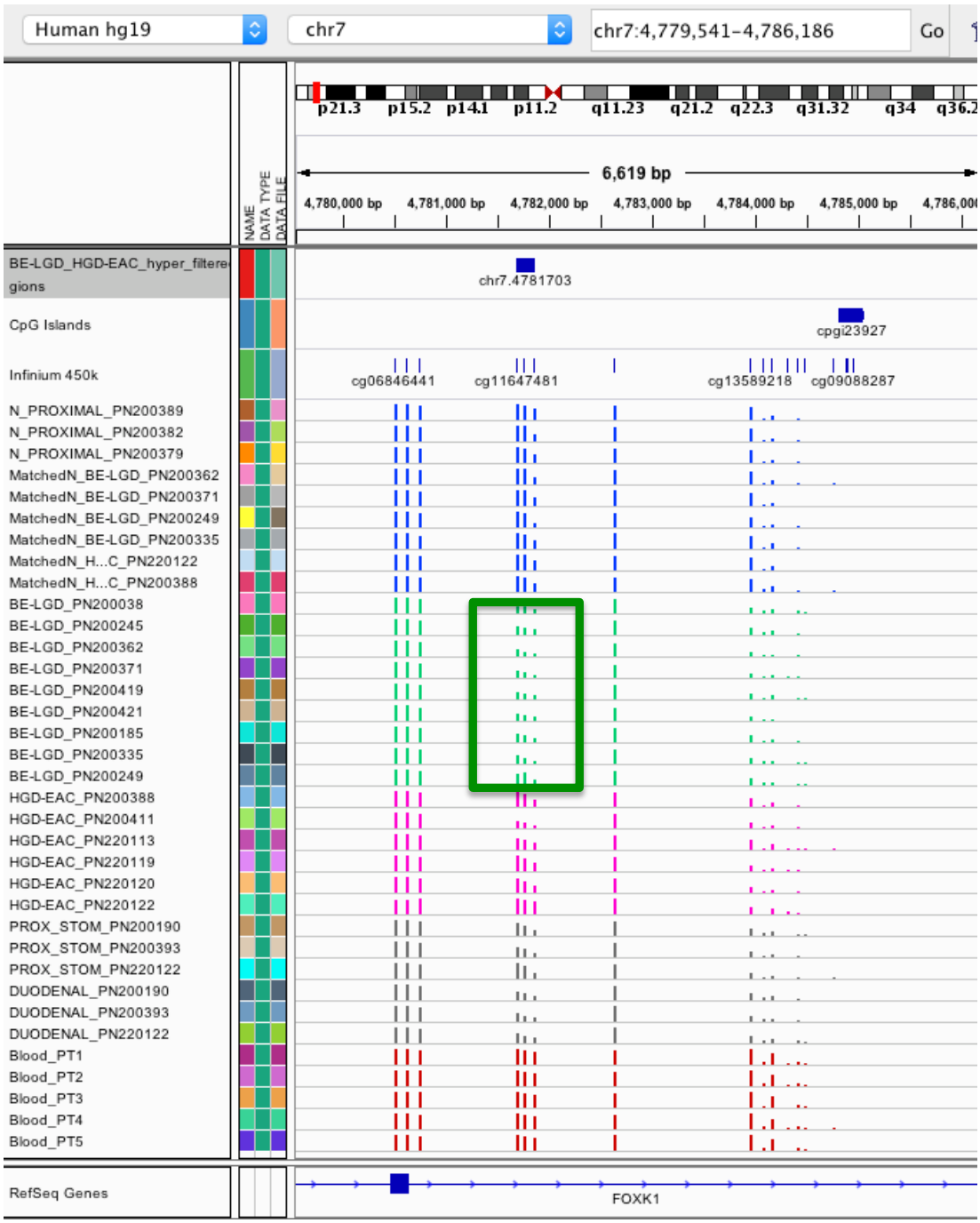
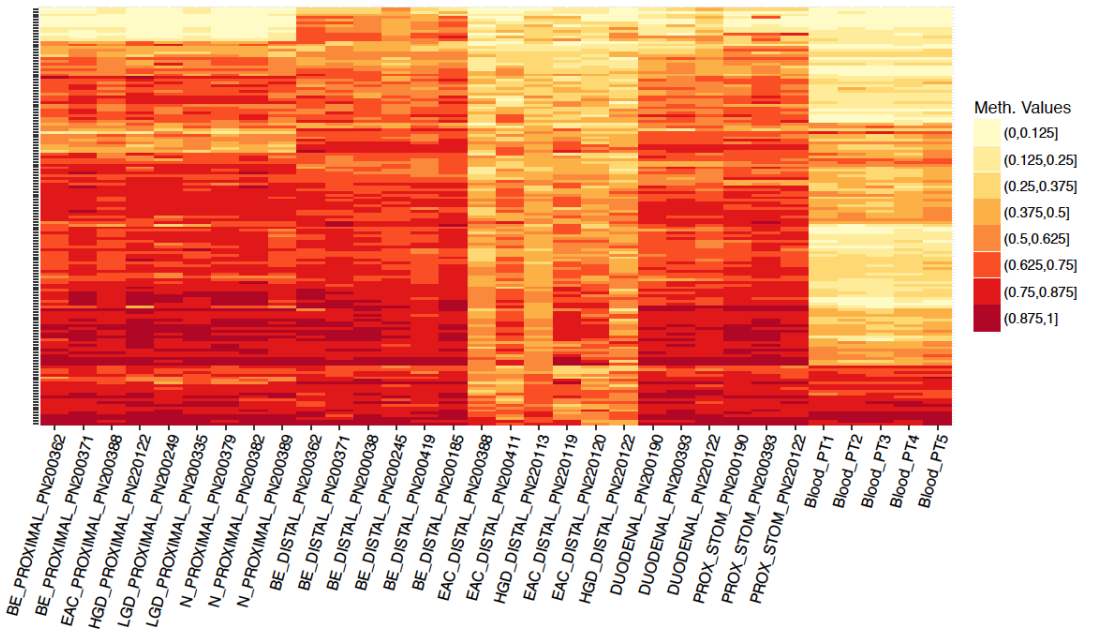


Figure 4-11: Forkhead box K1 (FOXK1) contains a region (chr7.4781703) highly methylated in all (esophageal adenocarcinoma tissues, normal esophageal mucosa and peripheral blood from normal healthy patients), but non-dysplastic Barrett's esophageal tissues (green box). Region with lower methylation in BE tissues is surrounded by regions of universally methylation across all disease types. Genome browser shows HM450 data for a subset

of the training cohort with >90% tissue sample homogeneity. Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

Results for hypomethylation in the BE v HGD-EAC comparison show high level methylation in normal esophageal mucosa and control tissues as well as non-dysplastic BE. The majority of HGD-EAC regions identified as hypomethylated still have mid-level methylation present. Methylation levels in blood for these targets range from absent, through mid-level and a subset of targets that are strongly methylated in blood (Figure 4-12(A)). As for other hypomethylation comparisons, these patterns are not suitable for biomarker selection and hence, no hypomethylation targets were carried through to validation and targeted sequencing studies. Clustering for this comparison shows distinct grouping of HGD-EAC and N samples with BE, duodenal and proximal stomach all maintaining distinct groups but cluster more closely to each other than N or HGD-EAC. (Figure 4-12(B)).

(A) BE v HGD-EAC Hypomethylation heat map



(B) BE v HGD-EAC Hypomethylation PCA plot

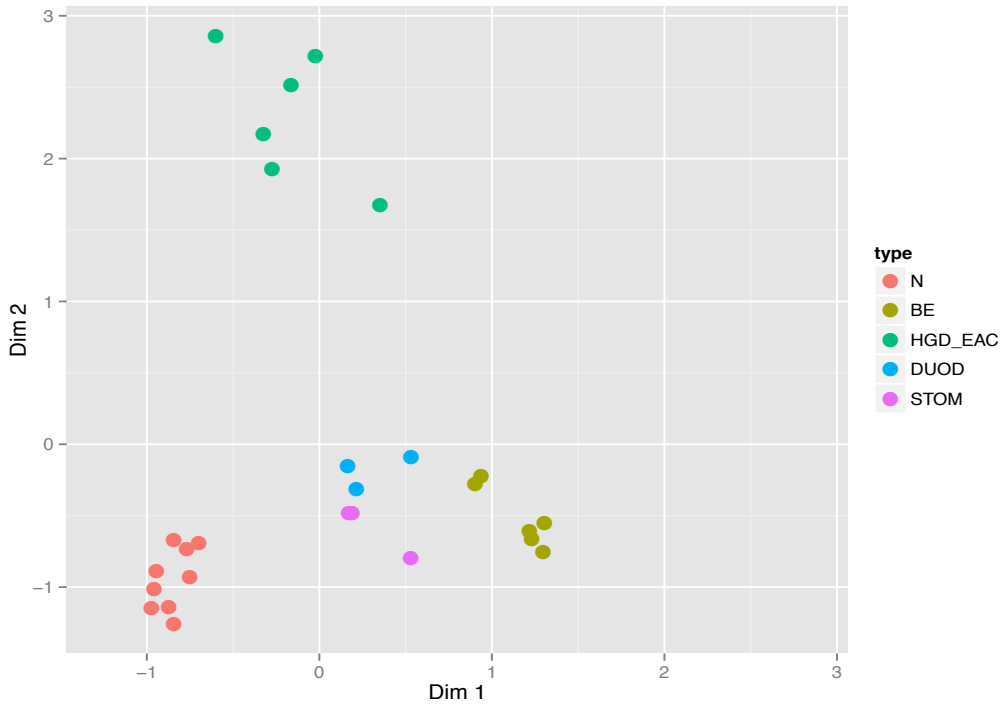


Figure 4-12: All hypomethylated CpG sites (n=153) detected in the BE v HGD-EAC comparison, FDR < 0.01 and Δβ ≥ 0.20 expressed as a (A) Heatmap showing patients (and their respective classification), with methylation values assigned as eight discrete regions from 0.000 – 1.000 in increments of 0.125 for better visualization (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

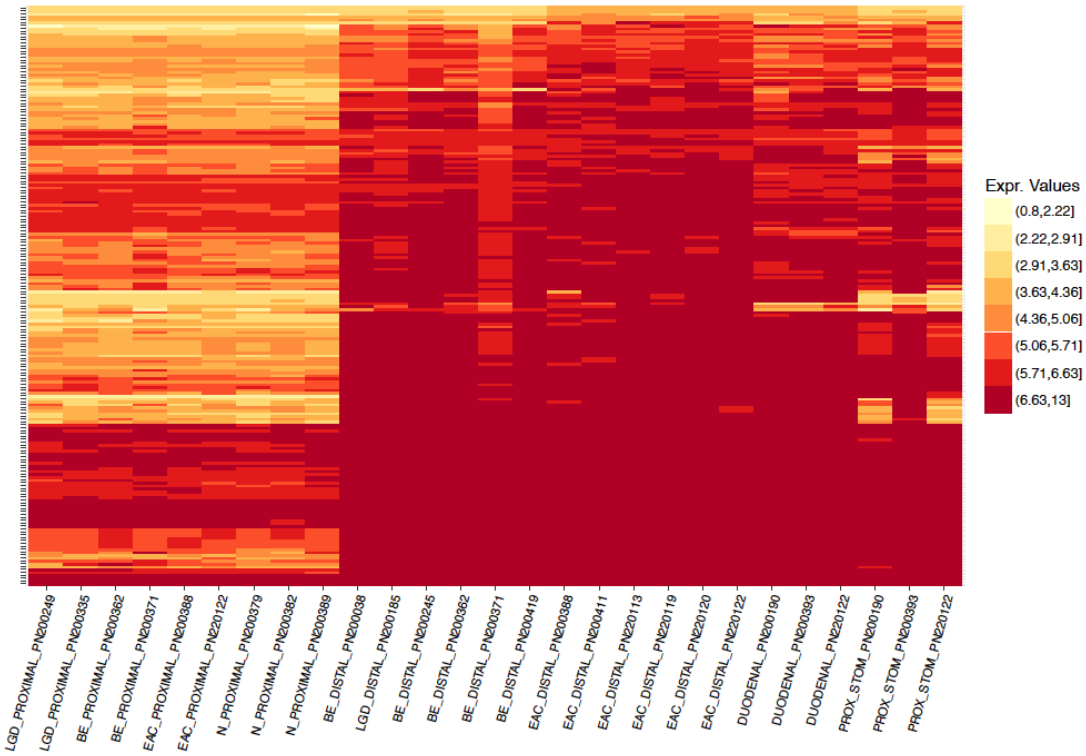
4.3.4.2 Genome-wide transcriptomic profiling

As for genome-wide methylation profiling, the largest differential expression differences were detected in HGD-EAC when compared to N (3,799 up-regulated (Figure 4-13) and 2,998 down-regulated (Figure 4-14) transcript clusters; adjusted p-value < 0.05 and logfc > log2(1.5)), followed by BE compared to N (2,512 up-regulated (Figure 4-15) and 1,984 down-regulated (Figure 4-16) transcript clusters; adjusted p-value < 0.05 and logfc > log2(1.5)) and finally BE compared to HGD-EAC (553 up-regulated (Figure 4-17) and 297 down-regulated (Figure 4-18) transcript clusters; adjusted p-value < 0.05 and logfc > log2(1.5)). Binning for heat maps was based on quantiles of expression values. Details of the top 100 up- and down-

regulated transcript clusters detected for each comparison group: N v BE, N v HGD-EAC and BE v HGD-EAC are given in Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling.

The most significantly up-regulated genes in HGD and EAC (with respect to normal esophageal epithelium) were, in the most part, also significantly up-regulated in non-dysplastic BE, duodenal and proximal stomach mucosa (Figure 4-13(A)). This pattern is also reflected in the tight clustering of normal esophageal mucosa samples and the general clustering of all other tissue types (BE, HGD, EAC, duodenal and proximal stomach) together (Figure 4-13(B)).

(A) N v HGD-EAC Up-regulated genes heat map



(B) N v HGD-EAC Up-regulated genes PCA plot

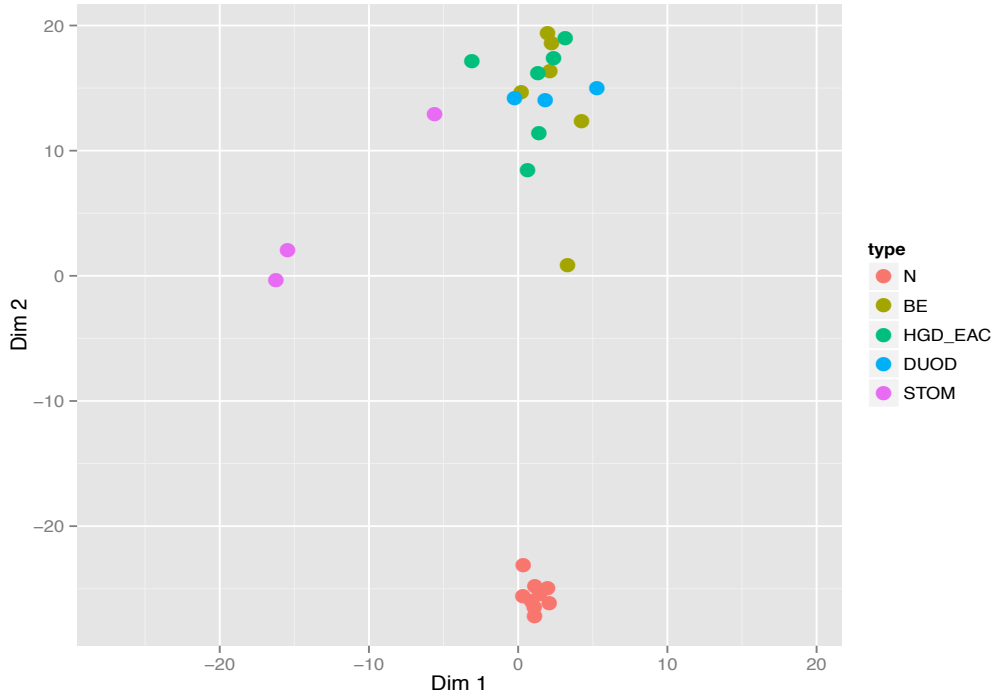
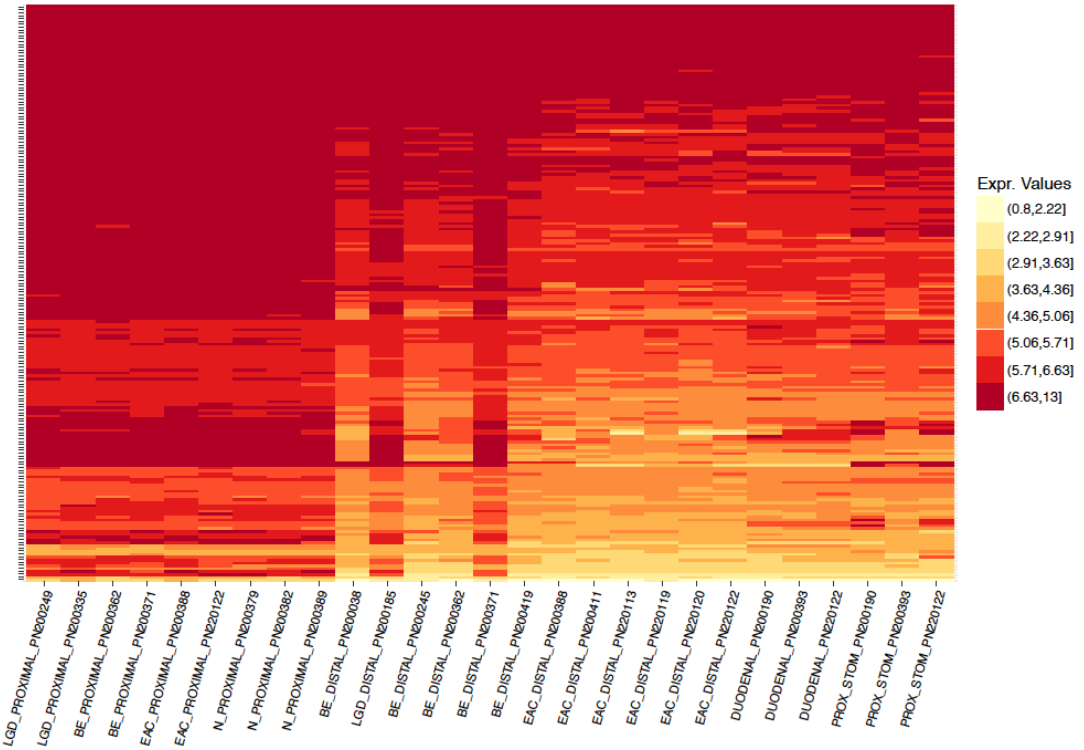


Figure 4-13: Top 200 up-regulated genes detected in the N v HGD-EAC comparison, FDR < 0.01 and FC > 1.5, expressed as a (A) Heatmap showing patients (and their respective classification), with expression values assigned into eight discrete regions ('bins') from 0.80 – 13.00 (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

The most significantly down-regulated genes in HGD and EAC (with respect to normal esophageal epithelium) are also down-regulated in non-dysplastic BE, as well as duodenal and proximal stomach control tissues (Figure 4-14(A)). For this comparison, normal esophageal mucosa samples cluster strongly, showing very similar global expression patterns with all other tissue types (BE, HGD, EAC, duodenal and proximal stomach) clustering together (Figure 4-14(B)).

(A) N v HGD-EAC Down-regulated genes heat map



(B) N v HGD-EAC Down-regulated genes PCA plot

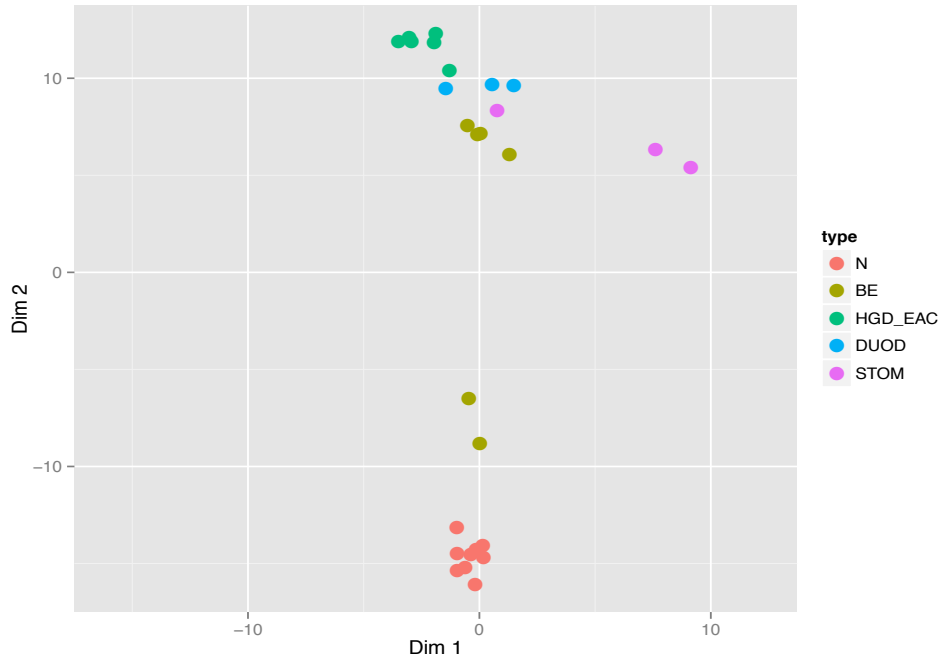


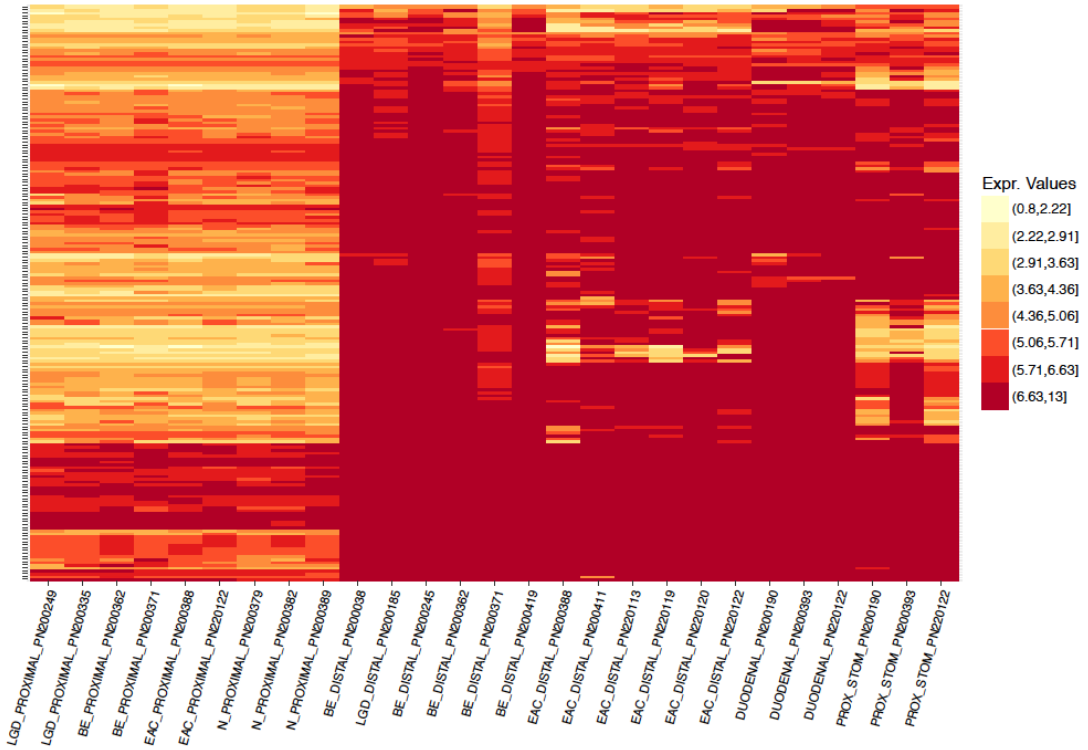
Figure 4-14: Top 200 down-regulated genes detected in the N v HGD-EAC comparison, FDR < 0.01 and FC > 1.5, expressed as a (A) Heatmap showing patients (and their respective classification), with expression values assigned into eight discrete regions (‘bins’) from 0.80 – 13.00 (B) Principal component

analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

From both the up- and down-regulated expression analysis of N v HGD-EAC, and the clustering of HGD-EAC with non-dysplastic BE and control tissues; we can hypothesize that much of the most significant differential expression is attributable to the maintenance of columnar mucosa rather than disease progression through the metaplasia-dysplasia-adenocarcinoma sequence. Thus we can conclude that tissue type difference (flat squamous cells comprising normal healthy esophageal mucosa as opposed to the columnar mucosa present in BE, HGD, EAC and control tissues) has a more significant impact on global expression profile than disease development.

Observations for up-regulation in N v BE are very similar to those for up-regulation in N v HGD-EAC. The most significantly up-regulated genes in BE (with respect to normal esophageal epithelium) were, in the most part, also significantly up-regulated in HGD and EAC, as well as duodenal and proximal stomach tissues (Figure 4-15(A)). As for the N v HGD-EAC comparison, tight clustering of the normal esophageal mucosa was separate from the general clustering of all other tissue types (BE, HGD, EAC, duodenal and proximal stomach)(Figure 4-15(B)).

(A) N v BE Up-regulated genes heat map



(B) N v BE Up-regulated genes PCA plot

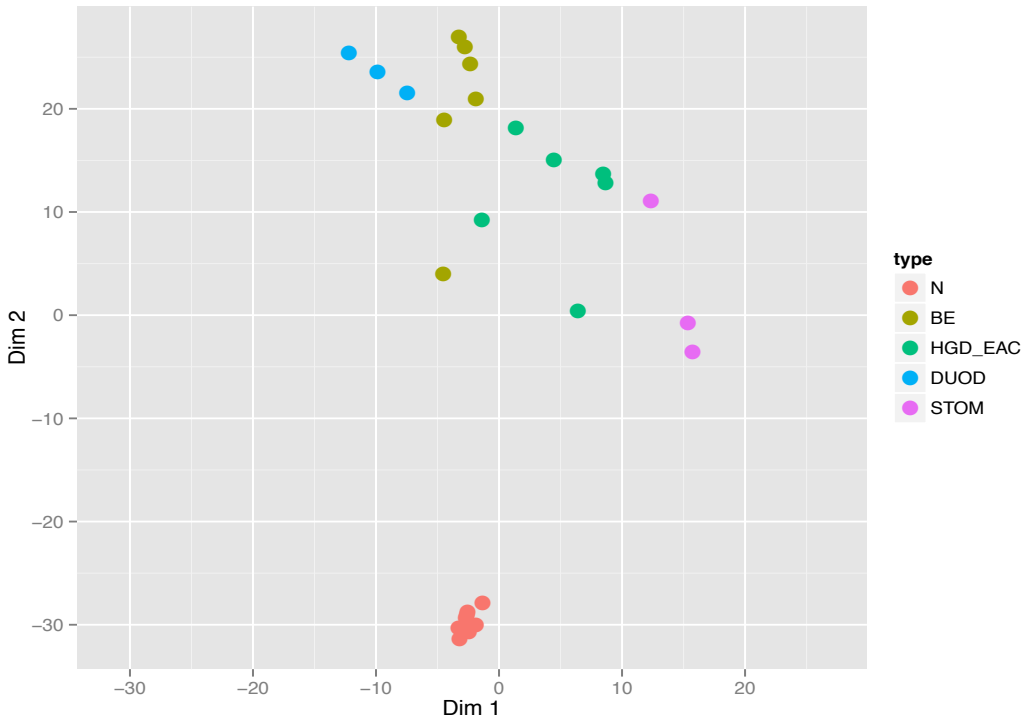
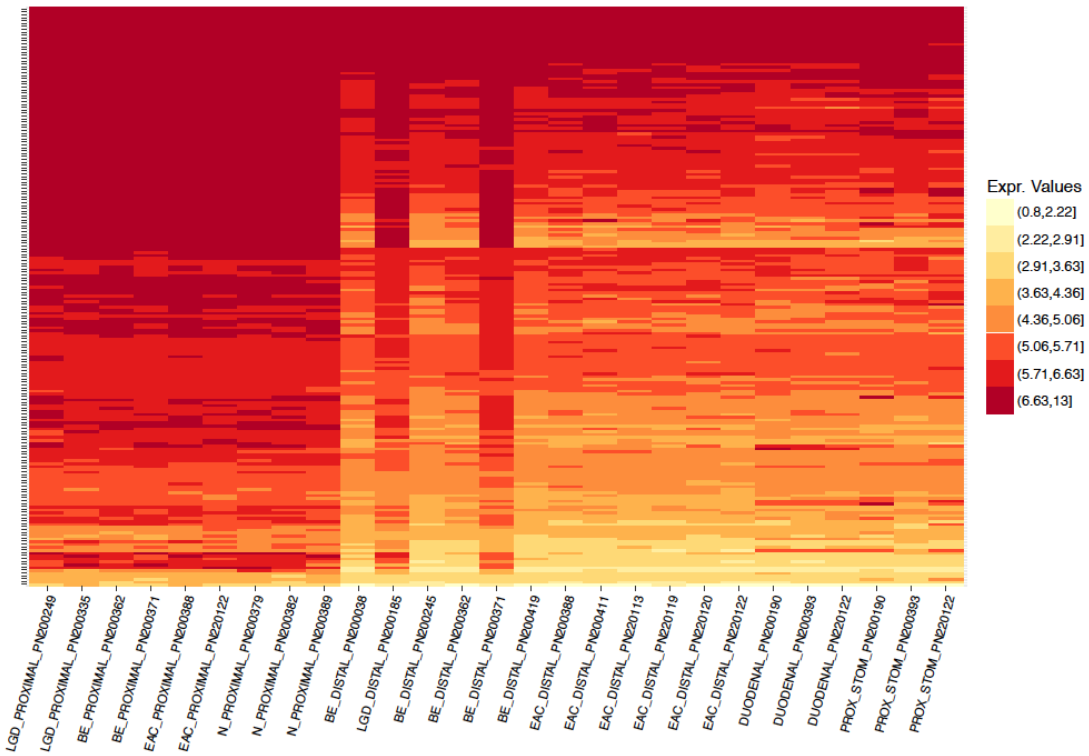


Figure 4-15: Top 200 up-regulated genes detected in the N v BE comparison, FDR < 0.01 and FC > 1.5, expressed as a (A) Heatmap showing patients (and their respective classification), with expression values assigned into

eight discrete regions ('bins') from 0.80 – 13.00 (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

Observations for down-regulation in N v BE were very similar to those for down-regulation in N v HGD-EAC with the most significantly down-regulated genes in BE (with respect to normal esophageal epithelium) also down-regulated in HGD and EAC, as well as duodenal and proximal stomach control tissues (Figure 4-16(A)). For this comparison, as for N v HGD-EAC, normal esophageal mucosa samples cluster strongly, showing very similar global expression patterns, with other tissue types (BE, HGD, EAC, duodenal and proximal stomach) clustering together (Figure 4-16(B)).

(A) N v BE Down-regulated genes heat map



(B) N v BE Down-regulated genes PCA plot

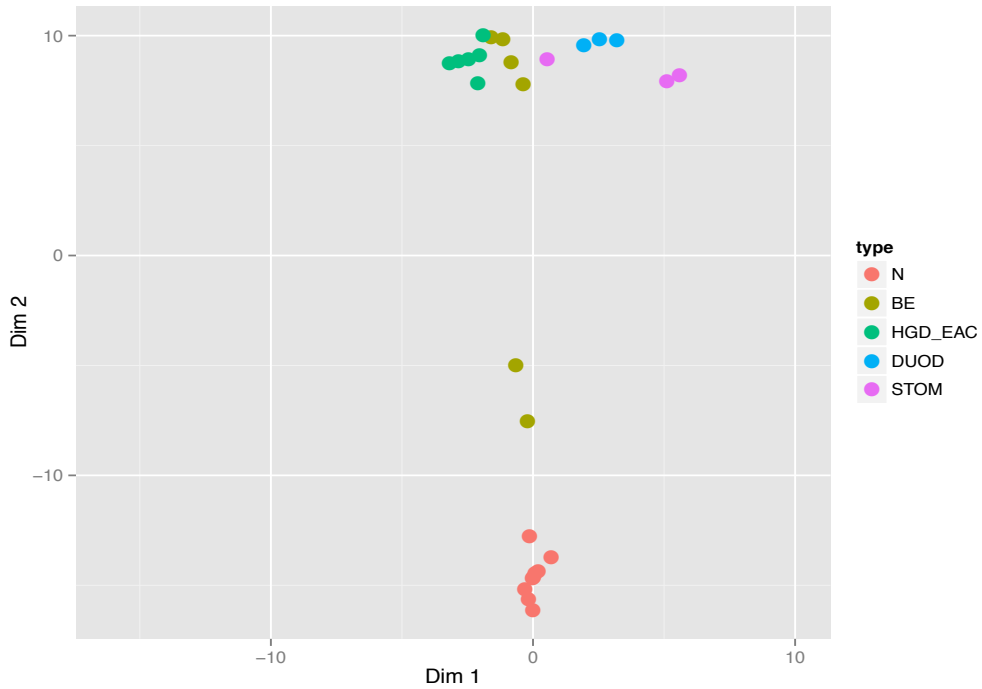


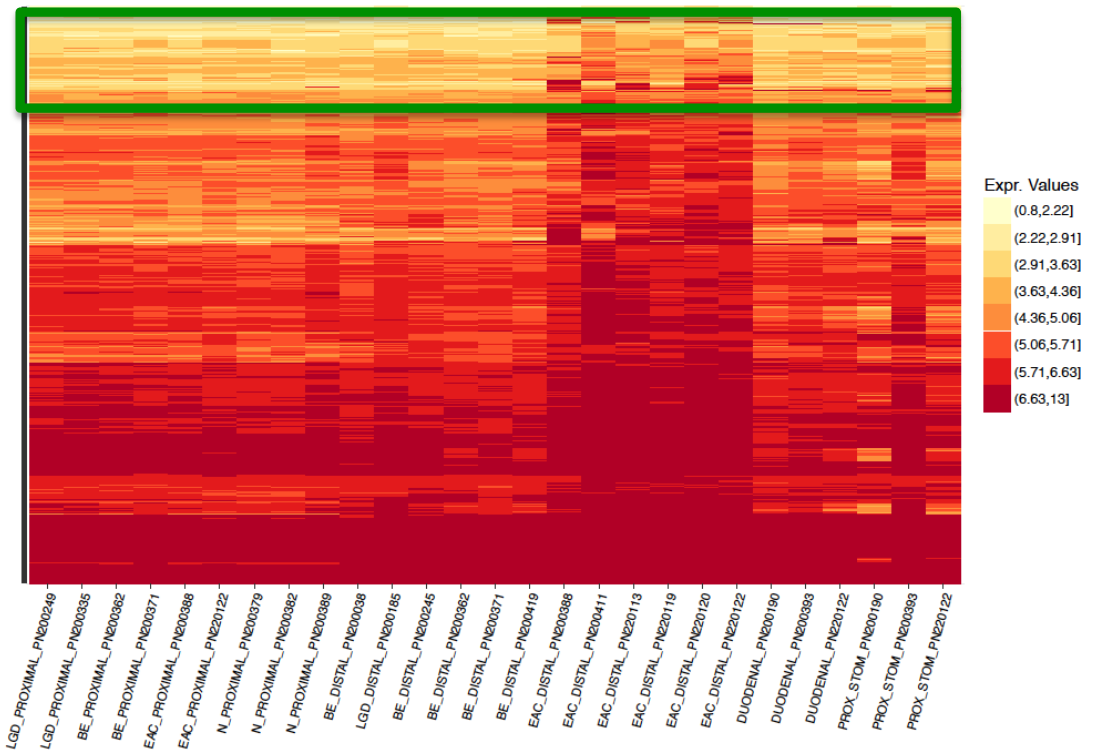
Figure 4-16: Top 200 down-regulated genes detected in the N v BE comparison, FDR < 0.01 and FC > 1.5, expressed as a (A) Heatmap showing patients (and their respective classification), with expression values assigned into eight discrete regions ('bins') from 0.80 – 13.00 (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

Thus the clustering and differential expression data from the N v BE comparison support our hypothesis that the most significant differential expression is attributable to the maintenance of columnar mucosa rather than disease progression through the metaplasia-dysplasia-adenocarcinoma sequence.

The most significantly up-regulated genes in HGD and EAC (with respect to non-dysplastic Barrett's epithelium) are not up-regulated in any other classification groups, meaning that this comparison is able to differentiate HGD and EAC not only from non-dysplastic BE, but also normal healthy esophageal mucosa, as well as duodenal and proximal stomach tissues (Figure 4-17(A)). Of clinical interest may be the subset of up-regulated HGD-

EAC samples with corresponding absence of, or very low level expression in all other classification groups (marked by a green box). Other identified differentially expressed genes in this comparison are biologically interesting, but not clinically valuable as biomarkers due to their need for defined expression level thresholds to differentiate intervention requiring disease from surveillance only patients. For this comparison, HGD-EAC samples cluster separately from all other tissue types (N, BE, duodenal and proximal stomach), however HGD-EAC clustering is not tight, indicating there is significant difference in global expression patterns between individuals with esophageal adenocarcinoma (Figure 4-17(B)). This supports increasing evidence that esophageal adenocarcinoma is a complex disease with different molecular subtypes[49], and ultimately, expression biomarkers may not be the best choice for a biomarker universally detectable across all these variable subtypes.

(A) BE v HGD-EAC Up-regulated genes heat map



(B) BE v HGD-EAC Up-regulated genes PCA plot

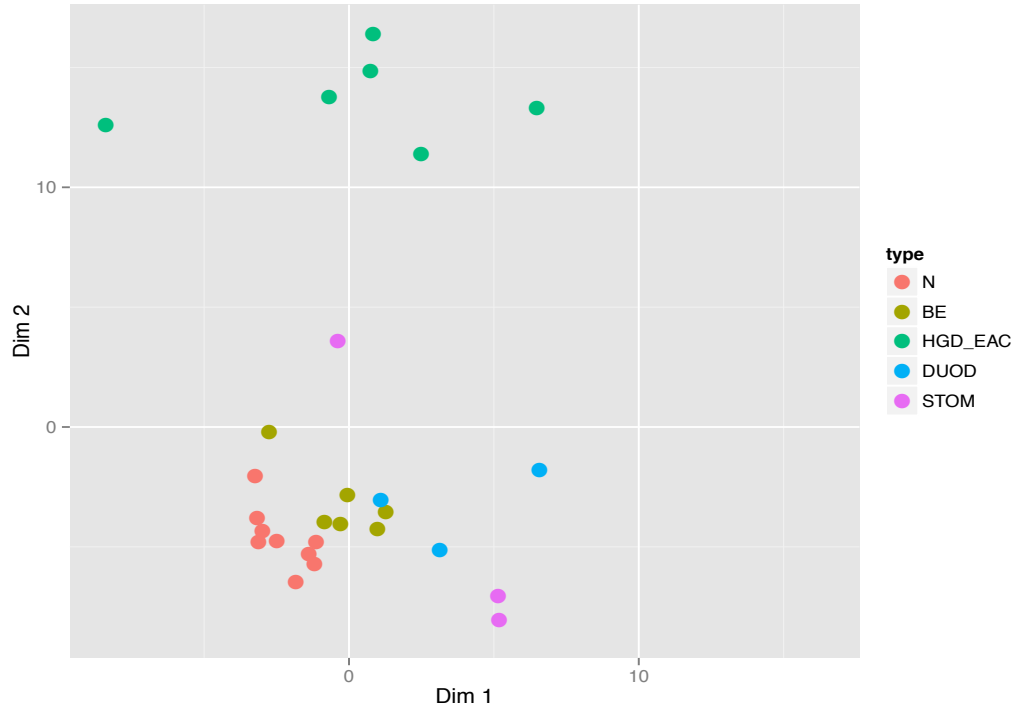
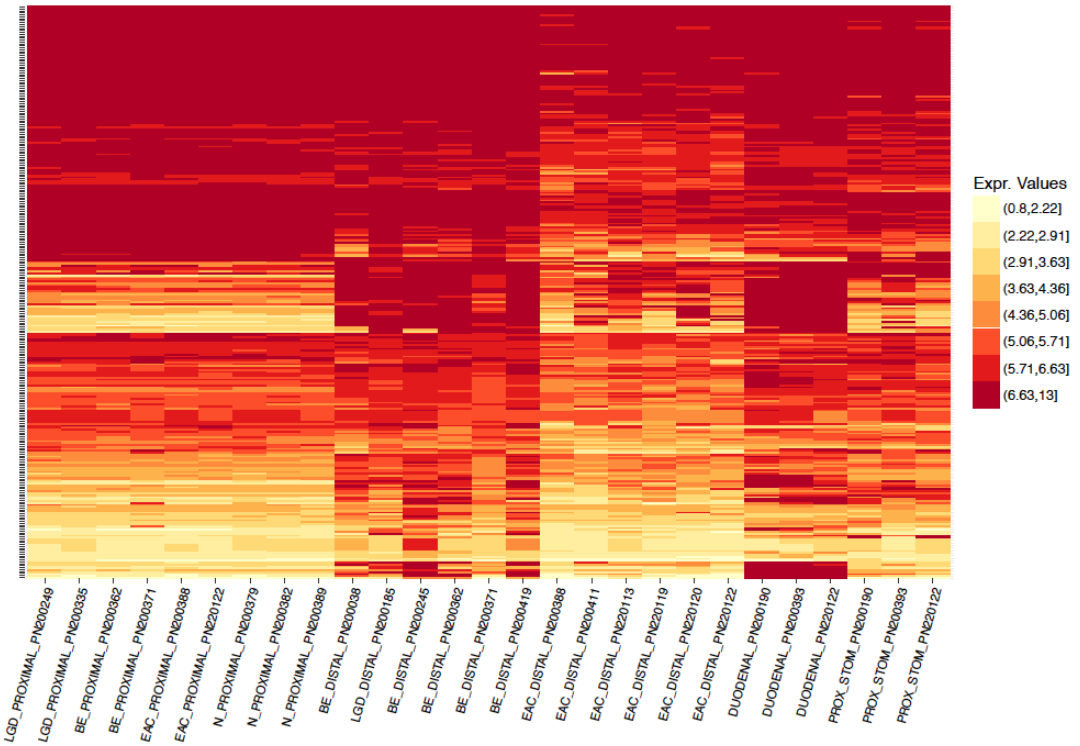


Figure 4-17: Top 200 up-regulated genes detected in the BE v HGD-EAC comparison, FDR < 0.01 and FC > 1.5, expressed as a (A) Heatmap showing patients (and their respective classification), with expression values assigned into eight discrete regions ('bins') from 0.80 – 13.00 (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

The most significantly down-regulated genes in HGD and EAC (with respect to non-dysplastic Barrett's epithelium) are also down-regulated in normal esophageal mucosa as well as proximal stomach tissues, but up-regulated in duodenal tissues (Figure 4-18(A)). These genes are not clinically useful as potential biomarkers due to an inability to differentiate between normal and carcinogenic tissues. The duodenum, or proximal small intestine is where most chemical digestion takes place and is structurally very similar to Barrett's mucosa. Both are columnar, goblet cell containing mucosae. Biologically, the genes identified as down-regulated in BE v HGD-EAC are likely not disease-associated (due to concurrent lack of expression in normal, healthy esophageal mucosa), but associated with maintenance and/or function of columnar mucosa. Interestingly, clustering for down-regulation in

BE v HGD-EAC results in separation of all different classification groups in just two dimensions (Figure 4-18(B)).

(A) BE v HGD-EAC Down-regulated genes heat map



(B) BE v HGD-EAC Down-regulated genes PCA plot

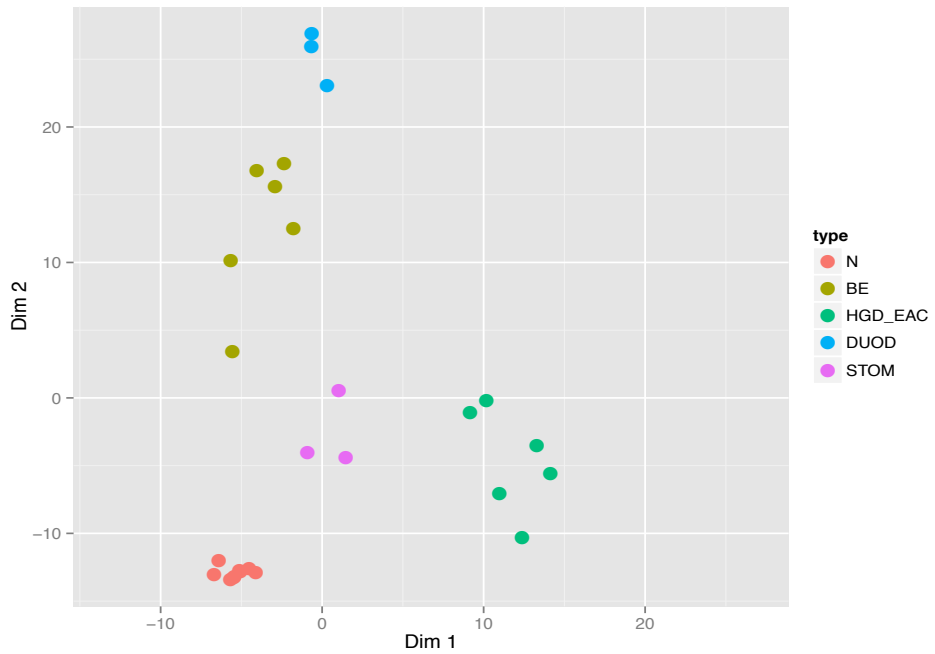


Figure 4-18: Top 200 down-regulated genes detected in the BE v HGD-EAC comparison, FDR < 0.01 and FC > 1.5, expressed as a (A) Heatmap showing

patients (and their respective classification), with expression values assigned into eight discrete regions ('bins') from 0.80 – 13.00 (B) Principal component analysis (PCA) plots, showing clustering of samples by classification (PC1 and PC2 only).

4.3.5 Integrative analysis: methylation and expression correlation

It has been generally accepted that cancer-associated promoter hypermethylation results in transcriptional silencing of tumor suppressor genes, whilst hypomethylation may lead to activation of normally inactive oncogenes. However, this overly simplistic methylation-expression correlation has recently come into focus as more and more evidence of a lack of correlation is reported from studies across various cancer types[114].

Correlation analysis of genome-wide methylation and expression profiling data did not support a strong relationship, especially for disease-associated, significantly differentially methylated regions. Methylation and expression correlation is examined in detail for selected disease-associated top differentially methylated target regions for biomarker validation in Section 5.3.8 Differential expression in selected target regions for validation.

Despite the overall lack of strong correlation, there was still evidence of individual cases supporting cancer-associated methylation-expression correlation. For example, one of the top identified hypo-methylated regions in BE v HGD-EAC (refer to Appendix 4, Table 3: Top 100 hyper- and hypo-methylated probes for differentiating intervention-requiring disease from Barrett's esophagus, 7th hypo-methylated entry (sorted by greatest $|\Delta\beta|$)) is BST2 (also known as tetherin or CS317 antigen). Bone Marrow Stromal Cell Antigen 2 (BST2), has been identified as an oncogene in a number of cancers, including breast cancer[115-118], head and neck cancer[119], lung cancer[120], myelomas[121, 122], glioblastoma[123], cervical cancer[124] and endometrial cancer[125, 126]. BST2 expression in human breast tumours and breast cancer cell lines has been reported to be inversely proportional to the methylation of CpG sites within, or in close proximity to the promoter region[115]. Epigenetic regulation of BST has also been reported in cervical

cancer[124], with hypomethylation resulting in activation of oncogene BST2 and resultant elevated expression. Here, we noted significant hypo-methylation in HGD and EAC samples with respect to non-dysplastic BE (Δβ -0.3643, p-value 4.01 $x10^{-5}$) as well as significant up-regulation in expression (2.22 fold change, p-value 0.0418).

Despite this widespread up-regulation of BST2 across numerous cancer types, there is also evidence of BST2 *down*-regulation in other solid tumors, and an unchanged expression profile in others (as outlined in Table 4-1), as outlined by Mahauad-Fernandez et al[127] in their 2016 review of the role of BST2 in host protection and disease manifestation.

Table 4-1: BST2 expression profile in various cancers. Reproduced from Mahauad-Fernandez et al, 2016[127].

	Profile of BST-2 transcript between normal and tumor tissue		
Cancer type	**Unchanged**	**Suppressed**	**Elevated**
Lung adenocarcinoma	X		
Thyroid cancer	X		
Lung squamous cell carcinoma		X	
Kidney papillary cell carcinoma		X	
Kidney chromophobe carcinoma		X	
Liver cancer		X	
Prostate cancer		X	
Head and neck cancer			X
Lung cancer			X
Breast cancer			X
Cervical cancer			X
Myelomas			X
Endometrial cancer			X
Glioblastoma			X

4.3.6 Gene set enrichment analysis

From Section 4.3.4 Global methylation and transcriptomic analyses, we hypothesized that much of the hypomethylation occurring across each of the comparison classes is attributable to change in mucosal structure (maintenance of columnar mucosa) rather than disease-associated progression through the metaplasia-dysplasia-adenocarcinoma sequence. Similarly for expression regulation, for example, the significantly down-regulated genes in both N vs HGD-EAC and N vs BE comparisons did not

differentiate between disease and control (duodenal and proximal stomach) tissues, as apparent in Figure 4-14(A) and Figure 4-16(A). Thus for gene set enrichment analyses (GSEA), the focus was placed on up-regulation occurring in Barrett's formation and esophageal carcinogenesis.

Analysis of up-regulation using MSigDB hallmark (H) collection (50 gene sets: representing specific, well-defined biological states or processes without overlap), computational (C4) gene sets (858 gene sets comprising 427 cancer gene neighborhoods[106, 128] and 431 cancer modules[129]) and oncogenic signatures (C6) (cellular pathway signatures often dis-regulated in cancer), revealed expected functional association as disease developed as well as highlighting the importance of examining molecular change between precursor disease and cancer, instead of the universally accepted 'normal vs tumor'.

Table 4-2 below shows hallmark gene sets up-regulated in HGD-EAC compared to normal healthy esophageal tissue. One of the greatest fold enrichments (approximately 4-fold) occurs in angiogenesis gene sets (see Figure 4-19). This is not surprising, as it is well known that solid tumors release angiogenic growth factors to stimulate blood vessel growth, providing vital nutrients and oxygen. Angiogenic factors have been linked with prognosis in esophageal adenocarcinoma patients but as yet there is no evidence for benefit from anti-angiogenic therapy[130, 131]. Other gene sets up-regulated in N v HGD-EAC include interferon alpha and interferon gamma response as well as inflammatory response and p53 pathway gene sets.

Table 4-2: Hallmark gene sets up-regulated in HGD-EAC compared to normal healthy esophageal tissue. Filtered p-value ≤ 0.05. List total (up-regulated N v HGD-EAC genes) = 316. Table sorted by fold enrichment (decreasing). pval: p-value, fold enr: fold enrichment, bonf: Bonferroni correction, benj: Benjamini false discovery rate adjustment, fdr: false discovery rate.

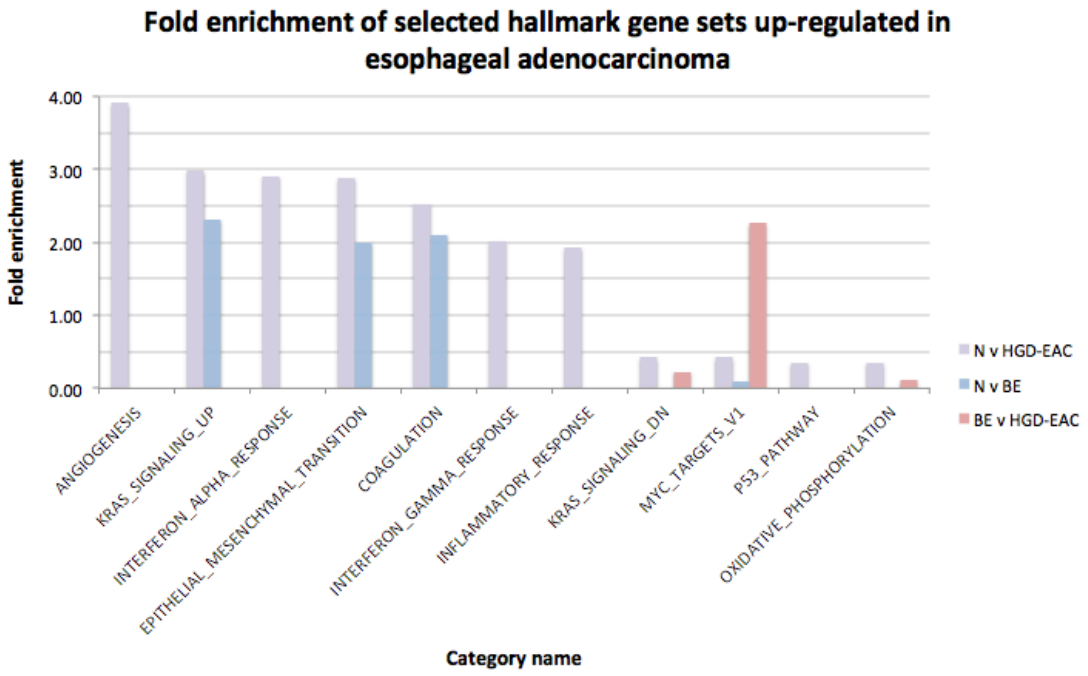
term	count	%	pval	genes	fold enr	bonf	benj	fdr
angiogenesis	10	3.16%	1.34E-04	CXCL6,FSTL1,ITGAV,LUM,NRP1, POSTN,SPP1,STC1,TIMP1,VCAN	3.91	6.71E-03	6.04E-03	1.34E-03

term	count	%	pval	genes	fold enr	bonf	benj	fdr
kras signaling up	43	13.61%	2.05E-11	ADAMDEC1,ANO1,ANXA10,BIRC3,C3AR1,CA2,CCL20,CFB,CLEC4A,CTSS,CXCR4,DUSP6,ELTD1,ETS1,FCER1G,GNG11,HKDC1,HSD11B1,IL1B,IL2RG,IL33,IL7R,INHBA,IRF8,MMD,MMP10,NRP1,PECAM1,PIGR,PLAUR,PLEK2,PSMB8,PTGS2,SCG5,SPP1,SPRY2,TFPI,TLR8,TPH1,TRIB2,TSPAN1,TSPAN13,USH1C	2.98	1.02E-09	4.60E-09	1.02E-09
interferon alpha response	20	6.33%	9.70E-06	BST2,DDX60,EPSTI1,IFI27,IFI30,IFI44,IFIT3,IFITM1,IFITM2,IFITM3,NMI,OASL,PARP14,PNPT1,PSMB8,PSMB9,RIPK2,RSAD2,RTP4,SELL	2.91	4.85E-04	5.84E-04	1.30E-04
epithelial mesenchymal transition	42	13.29%	1.14E-10	ANPEP,CALU,CDH11,COL12A1,COL6A3,CXCL1,CXCL6,DAB2,DCN,FAP,FMOD,FSTL1,HTRA1,IL8,INHBA,ITGAV,LAMA3,LAMC1,LAMC2,LGALS1,LUM,MATN2,MEST,MGP,MMP1,MMP2,MMP3,NT5E,PLAUR,PMP22,POSTN,SGCB,SPARC,SPP1,TFPI2,THBS1,TIMP1,TNFRSF11B,TPM1,TPM2,VCAM1,VCAN	2.88	5.68E-09	1.28E-08	2.84E-09
coagulation	25	7.91%	1.04E-05	A2M,ANG,C2,CFB,CFI,CTSE,CTSK,DPP4,DUSP6,F2RL2,GDA,HMGCS2,HTRA1,MEP1A,MMP1,MMP10,MMP2,MMP3,MMP7,PECAM1,SERPINA1,SPARC,TFPI2,THBS1,TIMP1	2.53	5.19E-04	5.84E-04	1.30E-04
interferon gamma response	29	9.18%	1.79E-04	BST2,CCL2,CFB,DDX60,EPSTI1,FPR1,HLA-B,HLA-DMA,IFI27,IFI30,IFI44,IFIT3,IFITM2,IFITM3,IRF8,METTL7B,NMI,OASL,PARP14,PNPT1,PSMB8,PSMB9,PTGS2,RIPK2,RSAD2,RTP4,SELP,TNFAIP6,VCAM1	2.01	8.94E-03	6.70E-03	1.49E-03
inflammatory response	28	8.86%	4.98E-04	AHR,AQP9,ATP2B1,BST2,C3AR1,CCL2,CCL20,CXCL6,FPR1,GPR183,HAS2,IFITM1,IL1B,IL7R,IL8,INHBA,LYN,MEP1A,MET,NMI,PLAUR,RIPK2,RTP4,SELE,SELL,SLC4A4,TIMP1,TNFAIP6	1.92	2.49E-02	1.60E-02	3.56E-03
uv response down	20	6.33%	3.39E-03	ANXA4,ATP2B1,CDC42BPA,DAB2,GCNT1,HAS2,ICA1,LAMC1,LTBP1,MET,NFIB,NRP1,PIK3R3,PLCB4,PMP22,PPARG,PRKCA,PTPRM,TFPI,TGFBR2	1.91	1.69E-01	7.62E-02	1.69E-02
complement	23	7.28%	1.50E-02	ANG,C2,C4BPB,CA2,CBLB,CFB,CTSL1,CTSS,CXCL1,DPP4,DUSP6,F5,FCER1G,GCA,LYN,MMP12,PLA2G7,PLAUR,PSMB9,SERPINA1,TFPI2,TIMP1,ZEB1	1.60	7.52E-01	2.42E-01	5.37E-02
il2 stat5 signaling	23	7.28%	1.78E-02	AHR,ANXA4,CA2,CDC6,ENPP1,F2RL2,GALM,GPR65,IFITM3,IL2RA,IRF8,ITGA6,ITGAV,NRP1,NT5E,ODC1,P4HA1,SELL,SELP,SLC2A3,SLC39A8,SPP1,SYT11	1.58	8.90E-01	2.67E-01	5.93E-02
g2m checkpoint	22	6.96%	3.03E-02	AURKA,BUB1,CASC5,CCNB2,CDC6,CDK1,CDK4,CDKN3,CKS1B,CKS2,HMMR,KIF15,MAD2L1,NEK2,ODC1,PBK,SLC12A2,SNRPD1,STIL,TOP2A,TPX2,ZAK	1.52	1.00E+00	4.25E-01	9.46E-02
estrogen response late	22	6.96%	3.34E-02	ABHD2,AGR2,ASS1,CA2,CDC6,DNAJC12,FARP1,HMGCS2,IL17RB,LAMC2,MEST,MYB,PDZK1,PLA2G16,SERPINA1,SLC27A2,STIL,TFF1,TFPI2,TMPRSS3,TOP2A,TSPAN13	1.50	1.00E+00	4.42E-01	9.83E-02
dna repair	5	1.58%	4.71E-02	EIF1B,NPR2,RAD51,RFC3,RFC4	0.48	1.00E+00	5.58E-01	1.24E-01
kras signaling down	6	1.90%	9.16E-03	CAPN9,FAM46C,HTR1D,PTPRJ,RSAD2,UGT2B17	0.42	4.58E-01	1.59E-01	3.52E-02
myc targets v1	6	1.90%	8.72E-03	CDK4,MAD2L1,ODC1,PSMB3,RFC4,SNRPD1	0.42	4.36E-01	1.59E-01	3.52E-02
myogenesis	6	1.90%	7.51E-03	COL6A3,FABP3,IGFBP7,NQO1,SPARC,TPM2	0.41	3.75E-01	1.54E-01	3.41E-02
p53 pathway	5	1.58%	2.77E-03	APAF1,DRAM1,GPX2,IFI30,SLC7A11	0.35	1.39E-01	6.93E-02	1.54E-02

Interestingly, a number of N v HGD-EAC gene sets also showed significant up-regulation in N v BE, including KRAS signaling, epithelial-mesenchymal transition (EMT) and coagulation gene sets (Figure 4-19). Despite reported

low KRAS mutational load in esophageal adenocarcinoma (supported by mutational screening performed as part of this study, refer to Section 7.3.2.2 BRAF and KRAS mutation in the esophagus), increases in KRAS expression have been reported in a subset of patients with poor prognosis[132]. KRAS signaling gene sets were significantly enriched in N v HGD-EAC (43 genes, 2.98 fold enrichment) and also N v BE (24 genes, 2.31 fold enrichment). In a recent 2016 paper, Kestens et al reported that inducing BMP4 signaling in both BE and EAC cells results in an invasive epithelial-mesenchymal transition-like phenotype[133]. In our data, we found EMT gene sets to be significantly enriched in N v HGD-EAC (42 genes, 2.88 fold enrichment) and also N v BE (21 genes, 2.00 fold enrichment). These data indicate that molecular change is occurring early in the metaplasia-dysplasia-adenocarcinoma sequence and may indicate a predisposition for malignant change in Barrett's patients.

(A)



(B)

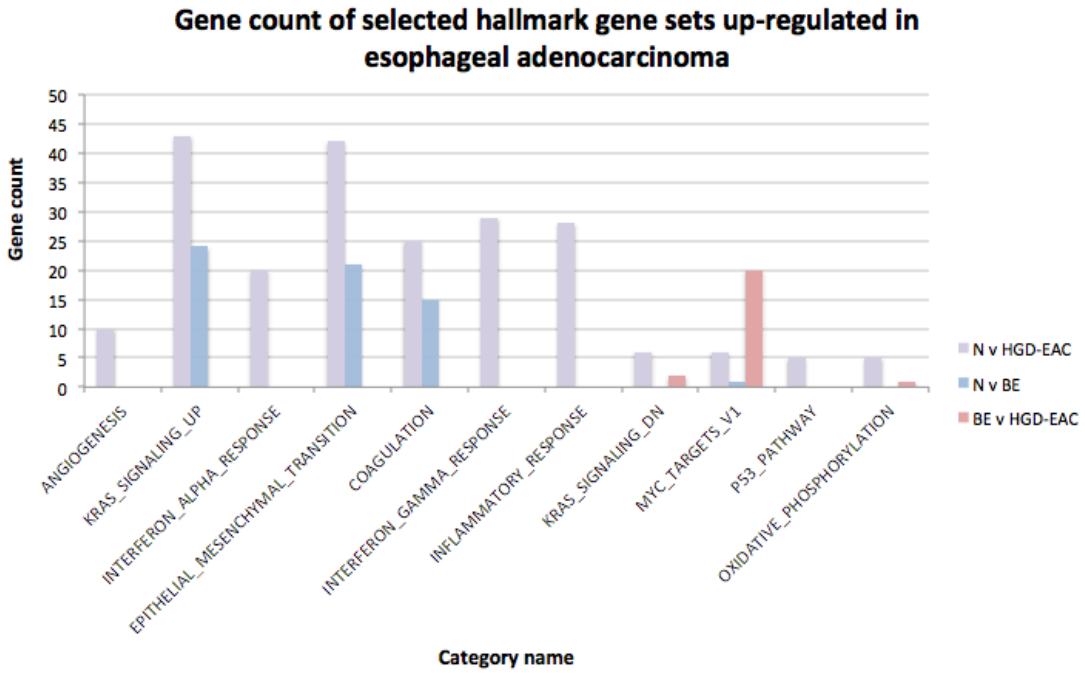


Figure 4-19: Selected hallmark gene sets up-regulated in HGD-EAC compared to normal healthy esophageal tissue. (A) Gene set fold enrichment for all comparison classes when statistically significant (p-value ≤ 0.05) (B) Gene count of up-regulated genes identified belonging to the specified gene set. The >40 KRAS signaling and epithelial-mesenchymal transition genes identified as up-regulated in HGD-EAC compared to normal, healthy esophageal epithelium each comprise approximately 13% of the full N v HGD-EAC up-regulated gene list (n=316).

Myc target genes are involved in cell growth, apoptosis and metabolism and have long been associated with oncogenesis in a variety of cancer types[134]. An interesting phenomenon is observed for the Myc targets gene set. Significant up-regulation of this gene set apparent in BE v HGD-EAC (20 genes, 2.26 fold enrichment), but not N v HGD-EAC (6 genes, 0.42 fold enrichment) (Figure 4-19). Furthermore, analysis of up-regulated computational gene sets (cancer gene neighborhoods (CGN)) in N v HGD-EAC reveals that the majority of these gene sets are actually showing a much more significant fold enrichment (and gene count hit) when examined in the BE v HGD-EAC gene list (Figure 4-20). Refinement of the comparison

results in enrichment for carcinogenic change. These observations highlight the importance of examining not only the commonly accepted normal vs tumor comparison, but investigating precursor disease vs tumor. It is especially important in the case of esophageal adenocarcinoma and biomarker discovery, as Barrett's esophagus is a benign and relatively common precursor and would create a very large unnecessary load on the health system should an early stage biomarker identify non-dysplastic precursor disease.

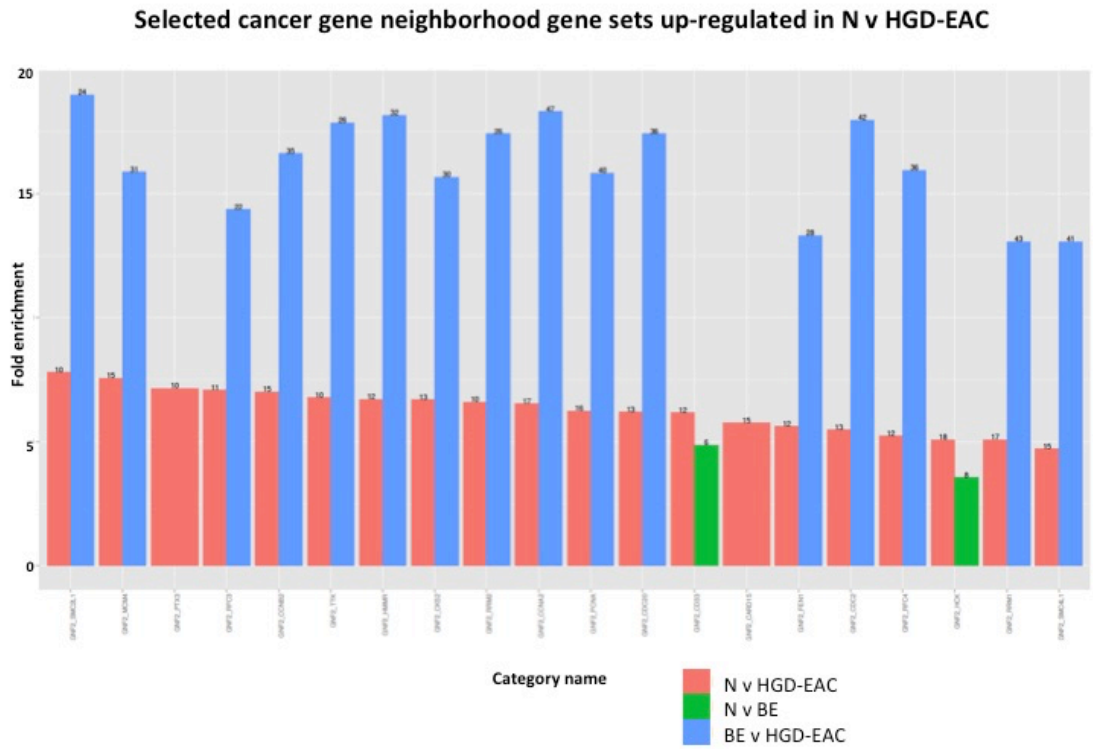


Figure 4-20: Selected cancer gene neighborhoods computational gene sets up-regulated in HGD-EAC compared to normal healthy esophageal tissue. Gene set fold enrichment for all comparison classes when statistically significant (p-value ≤ 0.05) reveals the greater fold enrichment of these genes when using BE v HGD-EAC comparison class. Gene count of up-regulated genes is given as a data label above the each bar.

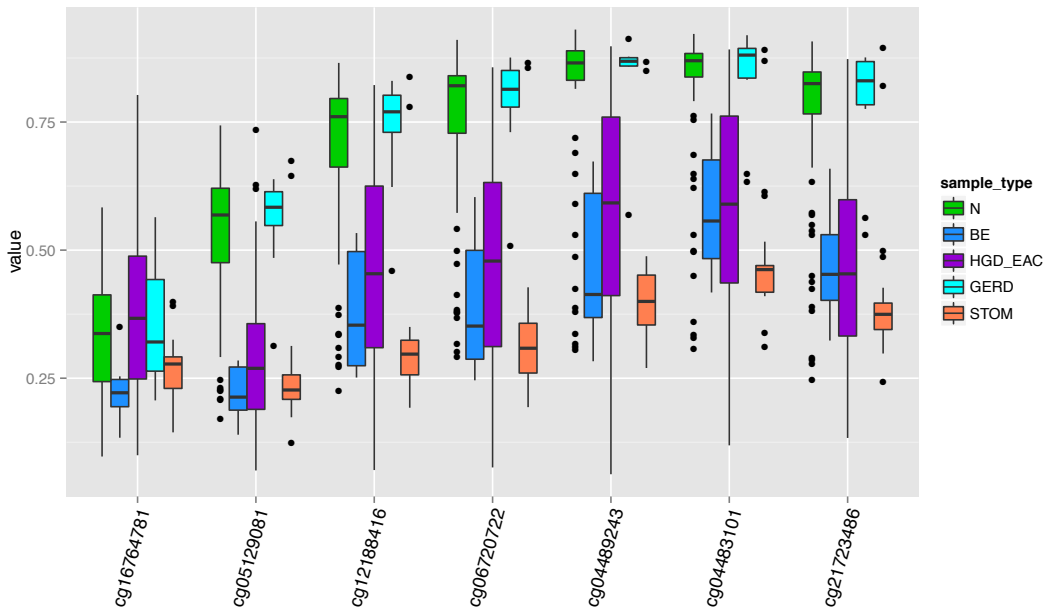
4.3.7 Methylation of tumor suppressor p63 in Barrett's carcinogenesis

Loss of tumor suppressor p63 expression has been associated with carcinogenesis in a variety of cancer types[102]. This, and the observation that

the p63 null mouse models quickly develop Barrett's-like tissue[104] led us to interrogate p63 promoter methylation in our samples. There are two promoter regions for tumor suppressor p63, which vary in disease-associated differential methylation (refer to Figures 4-19 and 4-20).

Promoter 1 shows increased methylation in N and GERD (GastroEsophageal Reflux Disease) and a subset of HGD-EAC tissues, with lower methylation observed in BE, proximal stomach and a subset of HGD-EAC tissues. Normal and GERD tissues become almost entirely methylated in the latter half of promoter 1 (Figure 4-21). However the inverse appears to be evident in promoter 2 (3 central CpG sites), where BE, HGD-EAC and stomach are almost entirely methylated, but only low levels of methylation observed in N and GERD tissues. Thus, we can hypothesize that increased methylation of tumor suppressor p63 promoter 2 may be a contributing factor to reduced p63 expression possibly implicated in development of Barrett's esophagus and/or carcinogenic progression.

(A)



(B)

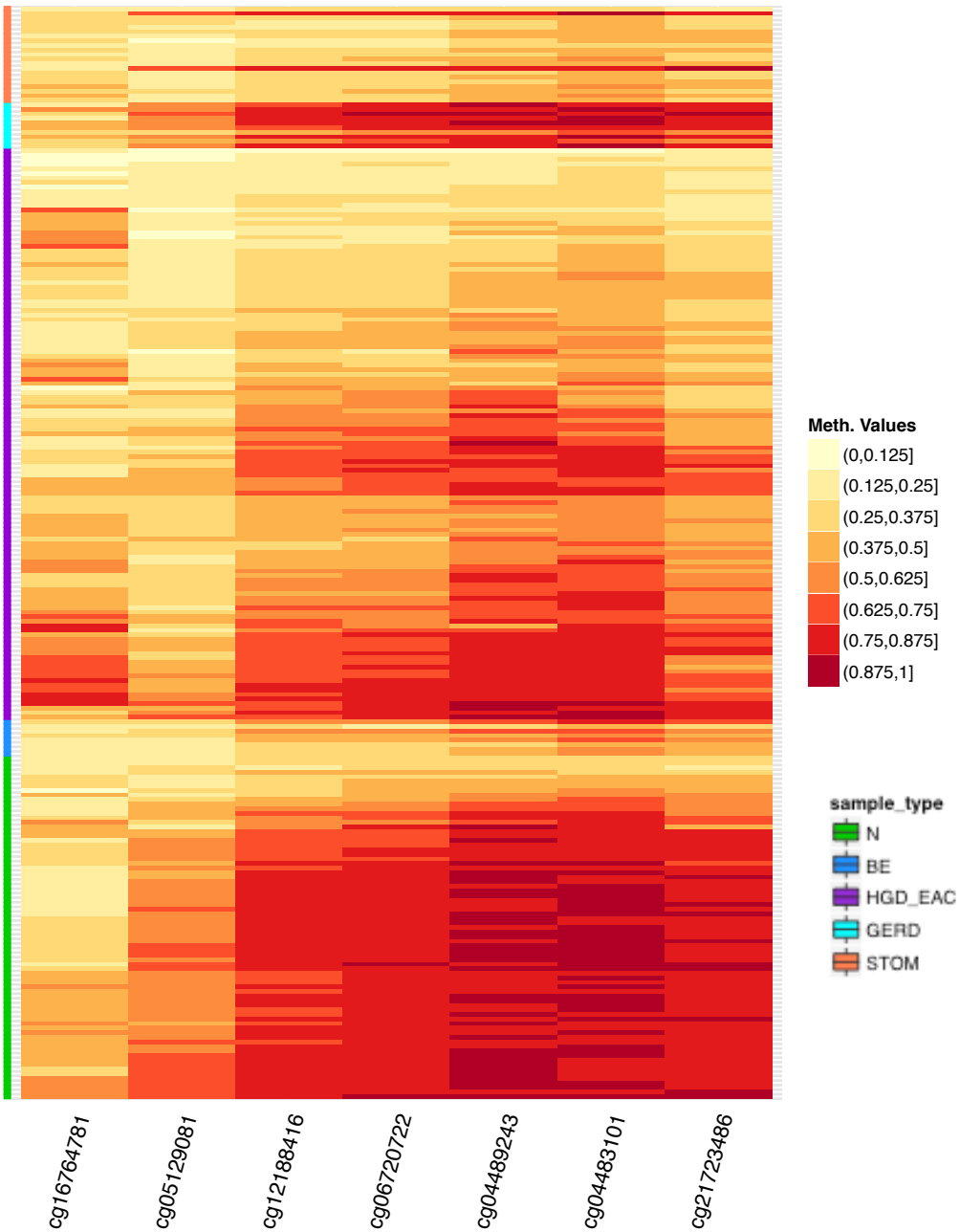
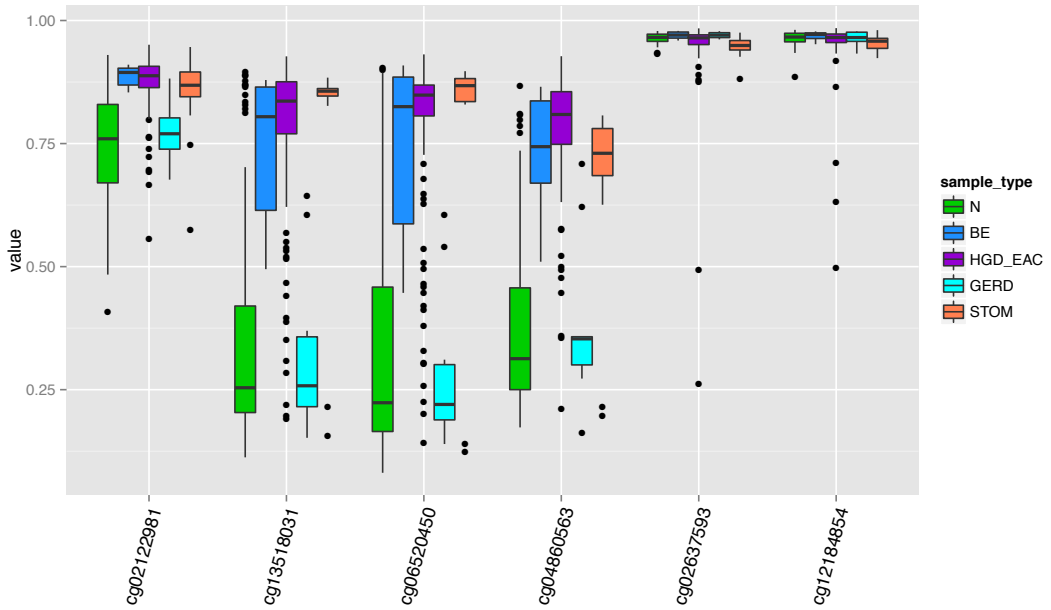


Figure 4-21: Average methylation at seven probes in p63 promoter 1 in the external validation cohort (N, BE, HGD-EAC, GERD and stomach tissues), expressed as a (A) Barplot (B) Heatmap (N in green, BE in blue, HGD-EAC in purple, GERD in light blue and stomach in orange), methylation binning into eight discrete regions from 0.000 – 1.000 in increments of 0.125.

(A)


1.00
0.75
0.50
0.25
value
cg02122981
cg13518031
cg06520450
cg04860563
cg02637593
cg12184854
sample_type
N
BE
HGD_EAC
GERD
STOM

(B)

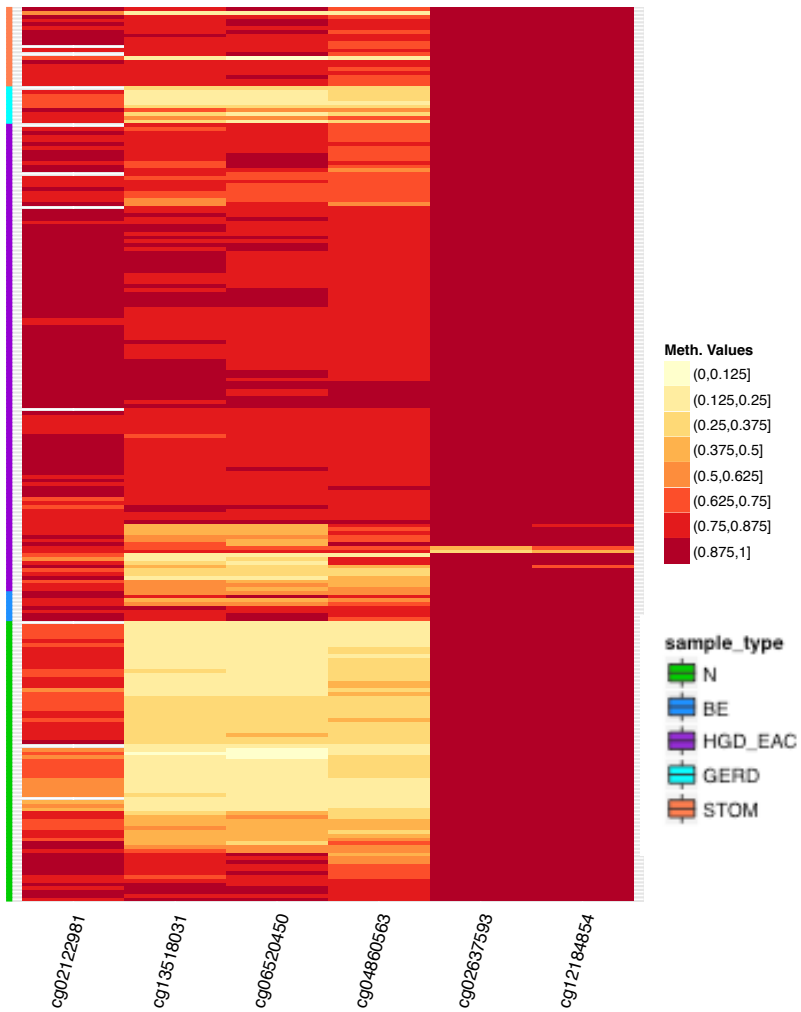

Meth. Values
(0,0.125]
(0.125,0.25]
(0.25,0.375]
(0.375,0.5]
(0.5,0.625]
(0.625,0.75]
(0.75,0.875]
(0.875,1]
sample_type
N
BE
HGD_EAC
GERD
STOM
cg02122981
cg13518031
cg06520450
cg04860563
cg02637593
cg12184854

Figure 4-22: Average methylation at seven probes in p63 promoter 2 in the external validation cohort (N, BE, HGD-EAC, GERD and stomach tissues), expressed as a (A) Barplot (B) Heatmap (N in green, BE in blue, HGD-EAC in purple, GERD in light blue and stomach in orange), methylation binning into eight discrete regions from 0.000 – 1.000 in increments of 0.125.

4.4 Discussion

The results presented in this chapter examine global methylation and expression change in the multi-phase development of esophageal adenocarcinoma. Gene set enrichment and analysis of DNA methylation-gene expression correlation are used to examine global molecular change at each phase of metaplasia-dysplasia-adenocarcinoma development. Finally, aberrant methylation of tumor suppressor gene p63 is examined in our data to understand its role in the origin and development of Barrett's esophagus and esophageal adenocarcinoma.

There was no significant difference in methylation profile of normal squamous epithelia, irrespective of sample biopsy location (distance aboral), as well as in proximal samples with any form of esophageal disease distally (non-dysplastic BE, dysplasia or esophageal adenocarcinoma). Hence matched normal samples (taken proximally from diseased patients) were used for further analyses; optimal as matched disease-normal pairs control for biological perturbations unique to an individual.

Global methylation analysis revealed significant methylation change between benign Barrett's and developed EAC, however more clinically significant was detectable methylation change between non-dysplastic BE and BE with low-grade dysplasia. Thus we can conclude that detection of aberrant methylation is viable for identification of initiation of carcinogenic progression at an easily treatable stage.

A downfall of the current endoscopic-histologic diagnosis method is difficult interpretation of small regions of destroyed glands. These regions may be

due to inflammation, tissue damage during histological preparation or more seriously, a focus of dysplastic or carcinogenic tissue. An interesting sample contained predominantly normal tissue (a mixture of normal squamous epithelia and cardiac mucosa) with a small section (approximately 5%) of destroyed glands, identified as focal dysplasia/adenocarcinoma by only 1 of 4 pathologists. Aberrant methylation was apparent and detectable globally in this sample, clinically important for more robust, reliable diagnosis of early stage disease in the future.

Differential methylation and expression were examined across three comparison classes, normal vs tumor, normal vs Barrett's and non-dysplastic Barrett's vs tumor. High-grade dysplastic samples were included alongside tumor samples.

The most significant hypermethylation detected in normal vs tumor comparison also showed significant hypermethylation in BE samples, highlighting that methylation change occurs early and is maintained in disease development but also highlighting deficiency in the comparison for identification of potential biomarkers. This is by no means universal for biomarkers across all cancer types, but due to the benign and common nature of Barrett's esophagus, as well as low incidence of carcinogenic progression, identification of non-dysplastic BE is undesirable for an effective biomarker. Using the Barrett's vs tumor comparison resulted in a smaller set of significant differential methylation, however actually enriches for disease-associated change. Using this comparison, a subset of regions was identified that are methylated in intervention-requiring disease (dysplasia and adenocarcinoma) but unmethylated in normal, stomach, intestinal and BE tissues, as well as in normal peripheral blood, important for potential biomarker development into a non-invasive blood test.

Analysis across all methylation comparisons revealed that much of the hypomethylation occurring is attributable to the maintenance of columnar mucosa, rather than disease-associated progression through the metaplasia-dysplasia-adenocarcinoma sequence. For this reason, no hypomethylation

targets were carried through to biomarker validation and targeted sequencing analysis.

The most significantly up-regulated genes in HGD and EAC (with respect to normal esophageal epithelium) were, in the most part, also significantly up-regulated in non-dysplastic BE, duodenal and proximal stomach mucosa

Genome-wide transcriptomic profiling in normal vs tumor as well as normal vs Barrett's comparisons showed agreement in the most significantly up-regulated genes. Furthermore, the majority of these genes were also up-regulated in control tissues, indicating that much of the increased expression is due to maintenance of columnar mucosa, rather than disease-associated. Similarly, this was found for down-regulation across all comparison classes. This analysis highlighted that tissue type difference (flat squamous cells comprising normal healthy esophageal mucosa as opposed to columnar mucosa present in BE, HGD, EAC and control tissues) has a more significant impact on global expression profile than disease development. Using the Barrett's vs tumor comparison enabled identification of a subset of genes with increased expression in only dysplasia and adenocarcinoma, however this was not a discrete, binary separation as observed for methylation; instead there was low level expression observed in normal, control and BE tissues. This analysis indicates potential utility of an RNA expression biomarker is compromised, as a cut-off must be set. DNA methylation appears to be a more ideal biomarker for early esophageal adenocarcinoma identification due to the large subset of results with binary 'on-off' results.

High quality genome-wide methylation and expression profiling data was obtained from the same esophageal tissue biopsies, enabling true examination of methylation-expression correlation. Our analysis showed evidence of generally accepted relationships: cancer-associated promoter methylation results in transcriptional silencing of tumor suppressor genes and hypomethylation leads to activation of normally inactive oncogenes. However correlation analysis of genome-wide methylation and expression profiling data did not support a strong relationship, especially for disease-associated,

significantly differentially methylated regions. A recent study by Moarii et al[94] hypothesized that cancer-specific methylation targets genes that are lowly expressed in normal tissues and may contribute to carcinogenesis by modifying expression susceptibility rather than simplistic direct inhibition. This, and evidence of lack of methylation-expression correlation in other cancers supports our findings for esophageal adenocarcinoma here[135-138].

Furthermore, there are limitations associated with use of microarray technology for genome-wide methylation profiling: most CpG sites assayed using HM450 are concentrated around promoter regions and gene bodies, however approximately 25% are located in intergenic regions[111], so not accompanied by gene association. Differential methylation in intergenic regions may be over transcriptional regulatory elements such as enhancers (involved in enhancer-promoter interactions), insulators, locus control regions or similar[139], and shouldn't be disregarded in favor of those with explicit gene association as potential biomarkers. Enhancer methylation, for example, can affect gene expression as far as 1Mb from transcription start site[140] and therefore correlation studies will not identify this associations. It was originally planned to use integrative analysis to prioritize disease-specific methylation change as potential biomarkers in the development of EAC. Based on our observed lack of strong methylation-expression correlation and recent publications challenging direct association[94, 135-138], emphasis for biomarker selection remained on significant disease-associated hypermethylation, as discussed further in Chapter 5.

Gene set enrichment analysis revealed expected functional association as disease developed, as well as highlighting the importance of precursor disease vs tumor over normal vs tumor comparisons for identification of disease-associated change. Enriched gene sets in esophageal adenocarcinoma development include those for angiogenesis, inflammatory response, cytokine (interferon alpha and interferon gamma) response as well as p53 pathways. Gene sets showing early enrichment in precursor disease as well as adenocarcinoma include KRAS signaling, epithelial-mesenchymal transition and coagulation. Examination of cancer-specific gene sets showed

the advantage of using Barrett's vs tumor comparison: significant fold enrichment (and gene count) was observed in these gene sets compared to using normal vs tumor gene list. For our purposes, using precursor disease vs tumor enriches for intervention-requiring disease-specific change, important for an early stage esophageal adenocarcinoma biomarker panel.

p63 is known to be involved in self-renewal of epithelial stem cells. The prevailing concept currently accepted for the origin of Barrett's is inflammation-triggered stem cell metaplasia at the gastro-esophageal junction. p63 is implicated in this process as null mice quickly develop BE like tissue[104]. Furthermore, p63 is a known tumor suppressor and loss of expression is associated with carcinogenesis in a variety of cancer types. Our data showed disease-specific methylation occurring in one of two p63 promoter regions, which may be a contributing factor to reduced p63 expression and its role in BE development and/or carcinogenic progression.

Overall, analysis of global differential methylation and expression change was informative about biological changes occurring with disease progression. It became apparent that aberrant hypermethylation has the most utility as a potentially clinically viable biomarker and the best comparison class to use to enrich for cancer-specific change is a benign precursor disease versus dysplastic disease-adenocarcinoma comparison. This was taken forward for target region validation and characterization, discussed in Chapter 5.

CHAPTER 5: TARGET REGION VALIDATION AND CHARACTERIZATION

5.1 Introduction

Complete and comprehensive validation of potential biomarker targets is an important phase of biomarker development following large-scale discovery studies. To ensure robustness and give biomarkers the best chance of successful downstream implementation to clinical use, it is important that different technology is utilized for validation and discovery studies. For example, a whole-genome microarray approach for discovery and targeted sequencing for validation. For discovery studies, a genome-wide approach is recommended, examining the entire genome in samples at all stages of carcinogenic progression. From here, focus can narrow to selected potential targets for validation. Technology such as MiSeq targeted amplicon sequencing is appropriate for validation studies as it is able to accurately quantitate methylation (or lack of) at each CpG site, at single-base resolution, across multiple regions of interest.

Both technical and independent validation is important; technical validation uses the same cohort (or a subset of) that used for discovery and ensures equivalent biomarker performance despite variations in methodology and analysis. An independent cohort ensures robustness of selected targets. Large sample cohorts are exceptionally valuable for biomarker validation, strengthening support for selected target regions as clinically valuable biomarkers before embarking on the next phase of development, usually retrospective longitudinal repository studies, as outlined in Pepe's landmark publication 'Phases of Biomarker Development for Early Detection of Cancer'[78].

Significant emphasis in the last decade has been placed on the development of non-invasive screening tests for early cancer detection, thus highlighting the importance of ability to detect biomarkers in biological specimens such as

stool, urine, saliva and blood. Cell-free circulating DNA from blood plasma has been found to carry methylation biomarkers for cancers occurring elsewhere in the body and are detectable at early stages of development [11, 22], providing the perfect vehicle for a non-invasive, early-stage cancer screening test. Thus to increase chances of clinical implementation, investigations into detection in blood should be carried out early in biomarker development studies.

In recent years, chromatin state annotation using computational segmentation models has emerged as a powerful tool for discovering regulatory regions and interpreting disease-association studies[141]. ChrommHMM is an automated computational system for learning chromatin states and can be used to characterize biological function of specific regions within the genome, for example if a region is an active transcription start site for a particular gene or a distal regulatory element able to influence transcription independently of its distance from a promoter[142]. Although this information does not influence biomarker selection or its likely success as a clinically valuable tool, it does provide valuable insight into mechanism by which the epigenetic regulation is influencing carcinogenic progression.

5.1.1 Chapter 5 aims

To establish, comprehensively validate and characterize disease-associated differentially methylated target regions indicative of early esophageal adenocarcinoma, concurrently initiating assay development with direct transferability to blood biomarker investigation.

5.2 Methods

5.2.1 Target region selection

Region selection for validation by targeted amplicon sequencing was based on cut-offs and exclusion criteria applied to genome-wide methylation profiling data, as outlined in Table 5-1. Furthermore, differential methylation (meeting cut-offs) was required to be present in at least two neighbouring (≤ 300bp apart) probes for the region to be considered for validation. Only

hypermethylated regions were considered for validation. The priority list was created by firstly applying a minimum baseline requirement, dictating baseline methylation cut-offs (β ≤ 0.05 (N comparisons) or β ≤ 0.10 (BE v HGD-EAC)). Secondly, minimum delta methylation (the change in methylation between baseline and disease) cut-offs were applied (Δβ ≥ 0.50 (N comparisons) or Δβ ≥ 0.20 (BE v HGD-EAC)). Thirdly, a series of exclusion criteria was applied to filter DMRs: removing regions with (i) methylation β > 0.10 in any normal peripheral blood sample (publicly available data set GEO accession GSE48472)[110], (ii) methylation β > 0.10 in any normal oesophageal tissue sample, (iii) methylation β > 0.10 in any control tissue sample (duodenal or proximal stomach tissue). If any probe within the region (regions defined as having at least 2 differentially methylated probes no further than 300bp apart) did not meet the specified criteria, the region was discarded or segmented to only include probes which met all selection criteria.

Table 5-1: Criteria and cut-offs for selection of regions for validation by targeted amplicon sequencing. Cut-offs apply to average methylation values detected for each probe within the defined multiple-probe target region.

	N v HGD-EAC	**N v BE**	**BE v HGD-EAC**
Baseline methylation (β)	≤ 0.05	≤ 0.05	≤ 0.10
Methylation change (Δβ)	≥ 0.50	≥ 0.50	≥ 0.20
Normal blood (β)	≤ 0.10	≤ 0.10	≤ 0.10
Normal esophageal tissue (β)	≤ 0.10	≤ 0.10	≤ 0.10
Duodenal tissue (β)	≤ 0.10	≤ 0.10	≤ 0.10
Proximal stomach tissue (β)	≤ 0.10	≤ 0.10	≤ 0.10

Exclusion criteria pertaining to proximal stomach and duodenal tissue controls ensures selected biomarker target regions are disease-associated rather than indicative of columnar mucosa maintenance. Exclusion criteria pertaining to human peripheral blood from normal, healthy individuals ensures applicability of selected target regions as blood biomarkers for early disease identification.

A priority ranked DMR list was created for each of the three comparisons, based on weighted ranking: 55% Δ methylation, 40% baseline methylation, 5% number of probes in region (0.55 Δβ + 0.40 baseline β + 0.05 probe count). An increased number of differentially methylated probes within the target region ensures robustness. Regions were shortlisted for selection with the most emphasis placed on validating regions from the clinically relevant BE v HGD-EAC comparison, for the identification of intervention requiring disease. Target regions from the overlap between N v BE and N v HGD-EAC (herein known as 'N v Any disease', identifying a subset of target regions that are methylated early and maintained in the metaplasia-dysplasia-adenocarcinoma sequence), were used to prioritize regions for validation from these two comparisons. Finally, differentially methylated regions / transcript pairs showing strong negative correlation were prioritised for assay design and optimisation.

As a result of identification and classification challenges surrounding low grade dysplasia as well as much lower levels of homogeneity within sample biopsies; primary training analysis did not include LGD data. However, whole genome methylation and expression profiling was performed on these samples and weight given to shortlisted training regions hypermethylated in LGD. LGD samples were included in cohorts for targeted sequencing validation.

5.2.2 Internal targeted amplicon sequencing validation

Targeted sequencing was used to interrogate the methylation status of individual CpG sites of selected regions in mucosa at all stages of EAC development, (i) performing technical validation of differentially methylated regions identified by genome-wide methylation profiling and (ii) further validating these regions in an independent cohort.

To perform targeted amplicon sequencing on selected regions:

1. Bisulphite-specific PCR (BSP) assays were designed to target regions and optimized, enabling unbiased amplification of both methylated and unmethylated DNA.

2. MiSeq library preparation was optimised for the specific set of target regions selected.
3. Amplification, library preparation and MiSeq analysis of both a technical and independent cohort was performed.
4. Targeted amplicon sequencing results were compared to genome-wide methylation profiling for validation of selected target regions.

5.2.3 Unbiased amplification of target regions

Bisulphite-specific PCR (BSP) assays were designed to selected target regions and optimized to enable unbiased amplification of both methylated and unmethylated DNA for targeted amplicon sequencing. DNA isolated from esophageal tissue biopsies was used for technical and independent validation of disease-associated aberrant methylation, however amplification assays were developed for easy transition to analysis of cfc-DNA extracted from human blood plasma. Primer/BSP assay design and optimization was based on protocols described by Clark et al 2006 in Nature Protocols[143].

5.2.3.1 Primer design

A region of interest was defined, containing the DMR, and up to 500bp either side; *only* if the methylation in the flanking regions reflected the differential methylation observed in the DMR. Common SNPs (present in ≥ 1% samples) within the region of interest were identified using UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly (dbSNP138 and dbSNP141). MethPrimer[144] was used to interrogate the region of interest sequence using the settings outlined in Table 5-2 below, attempting optimal settings, then the relaxed settings if this failed.

Table 5-2: General parameters for BSP primer selection using MethPrimer[144]. Optimal settings were used as a starting point, then relaxed if no primers were found.

	Optimal Settings			**Relaxed settings**		
Product CpGs	5			3		
Primer non-CpG C's	4			3		
Primer poly-X	5			5		
Primer poly-T	8			8		
	Min	**Optimal**	**Max**	**Min**	**Optimal**	**Max**

	Optimal Settings			Relaxed settings		
Product Size (bp)	50	100	250	50	100	300
Primer Tm (°C)	55	60	65	50	60	65
Primer size (bp)	22	28	32	20	28	35

Due to inherent high CpG content (and hence resultant high poly-T following bisulfite conversion), MethPrimer was often unable to return any results, even when using relaxed settings. Primers were designed manually in this case, using the following guidelines:

- If possible, primers were positioned over non-CpG containing sequence (if this was unavoidable, primers were placed over regions with one CpG site, for which a degenerate base was used). Degenerate bases were positioned as far from the 3' end as possible (Y for C in forward primer, R for G in reverse primer).
- Primers were positioned so as not to contain common SNPs of any type. Non-CpG SNPs were allowed within the amplicon, however CpG SNPs were not allowed at all within the defined amplicon.
- Primer pairs were melt temperature (Tm) matched to within 1.0°C. Calculated Tm for primers can vary significantly, depending on the calculation method used. Tm was checked with Primer3Plus[145] as well as ThermoFisher Scientific Multiple Primer Analyzer, which uses a modified nearest-neighbor method based on Breslauer et al[146].
- Primers were positioned to have one or more T's from non-CpG C's at or near 3' end, ensuring amplification of bis-modified DNA only.
- Primers were checked for minimal predicted self-dimer and cross primer-dimer formation using ThermoFisher Scientific's Multiple Primer Analyzer.
- Amplicons were designed to be 100-150bp in length; enforcing a 70bp minimum (<70bp problematic for library preparation) and a 200bp maximum (>200bp not ideal for investigation of fragmented cfc-DNA).

If possible, BSP assays were positioned within or containing the identified DMR, otherwise within the flanking region of interest. Up to 5 primer pairs on DNA top and opposite strands were designed and the best three ordered (judged on predicted self-/cross-primer dimer formation, predicted secondary

structure formation, closest match in Tm and most optimal amplicon length (100-150bp)).

Primer oligonucleotides were ordered from Integrated DNA Technologies (IDT), lyophilized at 25nmole with standard desalting. Nuclease-free water was added to create a 100µM stock solution of each oligonucleotide, and a 1:1 serial dilution performed to obtain 100µL of a 50µM working stock. Both the stock and the working solutions were stored at -20°C.

5.2.3.2 Standard material for amplification optimization

Universally methylated (Millipore CpGenome Universally Methylated DNA (Cat #S7821)) and universally unmethylated (Roche Human Genomic DNA from blood (Cat #11 691 112 001) DNA was used as standard material for BSP assay optimization. Roche human blood unmethylated standard material was implemented following testing (originally enzymatically modified DNA Set from Zymo Research (Human HCT116 DKO Methylated and Non-methylated DNA set (Cat #D5014)); and doubles as an additional control, ensuring the regions of interest are unmethylated in normal human blood.

Bisulfite-converted DNA was prepared at a range of compositions: 100% unmethylated DNA (100U), 100% methylated DNA (100M), 50% unmethylated + 50% methylated DNA (both a pre- (50:50 Pre) and post- (50:50 Post) bisulfite conversion mixture. A 'wild-type' control (WT, 50:50 mixture of non-bisulfite converted methylated:unmethylated DNA), was prepared as a control to ensure assays amplify bis-converted DNA only.

DNA standard material was lysed (Section 2.2.5.1 DNA lysis) to assist in achieving maximal bisulfite-conversion efficiency and in keeping with later treatment of clinical specimens, where protein may still be associated with the DNA; then bisulfite converted (as per Section 2.2.5.2 Bisulfite conversion of DNA). Following quantification (Section 2.2.5.3 Quantification of bisulfite converted DNA), all DNA standard material was diluted to 2ng/µL in nuclease-free water (Qiagen Nuclease-free water, Cat #129114) prior to storage at -20°C.

5.2.3.3 BSP optimization

BSP assays were optimized to ensure equal, unbiased amplification of both methylated and unmethylated DNA input. This was done using duplicate inputs of DNA standard material: 100U, 50:50 Pre, 50:50 Post, 100M, WT-control and NTC (nuclease-free water in lieu of DNA template). BSP assays were performed as outlined in Section 2.2.5.4 Bisulfite-specific PCR (BSP). Variation in $MgCl_2$ concentration (1.0mM – 2.5mM, in increments of 0.5mM) and anneal temperature (1°C increments) was used to adjust bias in amplification of methylated and unmethylated DNA, evaluated by heat dissociation melt curve analysis (as described in Section 2.2.5.5 DNA methylation detection). Additional cycling (up to 5 cycles) was performed (target dependent) if required.

For all assays deemed to be optimized by heat dissociation melt curve analysis, PCR product (one lane each from amplification of 100U, 50:50 Pre, 50:50 Post, 100M, WT-control and NTC) was run on a 2.5% agarose gel (as per Section 2.2.2 Agarose gel electrophoresis) to confirm amplification of a single product of the expected size, free from non-specific product formation.

5.2.4 MiSeq targeted sequencing

Protocol for MiSeq targeted amplicon sequencing includes:

1) Preparation of pooled amplicons
2) MiSeq library preparation and quality control
3) Quantification of PCR-competent sequencing template
4) MiSeq targeted sequencing run

5.2.4.1 Preparation of pooled amplicons

DNA extracted from esophageal tissue (technical and independent validation cohorts) was adjusted to 27.78ng/μL with nuclease-free water, taking 500ng (18μL) for each sample. For the independent validation cohort (amplification of 18 targets), each sample was prepared in duplicate.

DNA was lysed and bisulfite-converted as per protocol outlined in Section 2.2.5 DNA methylation studies, then diluted in nuclease free water, based on the number of target regions being amplified (quadruplicate 2.5µL input per target region required). For technical validation (10 target regions), DNA was made up to 120µL (included budgeted excess). For independent validation (18 target regions, duplicate sample preparations), pooled eluted DNA was made up to 240µL; resulting in DNA input of 7.5-10ng per PCR reaction (based on 70-100% bisulfite conversion yield). Diluted DNA was aliquoted into 11.8µL fractions prior to storage at -20°C.

PCR amplification of target regions was performed as per protocol in Section 2.2.5.4 Bisulfite-specific PCR (BSP), excluding SYTO® 9 addition and the associated melt curve cycling program. Technical (10 target regions) and independent (18 target regions) cohorts consisted of 24 samples each, amplified in quadruplicate in a 96-well plate (one full plate per target region). Quadruplicates pooled prior to storage at -20°C. Three example samples per amplicon (one representative sample per classification: N, BE, HGD/EAC) were quantified using the broad range Qubit® kit, as outlined in Section 2.2.4 Qubit® fluorometric double-stranded DNA quantification. For each patient sample, 150-200ng (where possible) of each amplified target region was pooled into a single tube as per Table 5-3.

Table 5-3: Amount of amplified target (ng) intended and actual, pooled per patient sample. Input amounts varied to compensate for sample loss due to amplicon size. Pooling for independent validation done in two batches (small (72-122bp) and large (125-196bp) amplicons). ^{ind}: regions analyzed in the independent cohort only. All other regions were analyzed in both technical and independent cohorts.

Target Region	Amplicon Size (bp)	Technical Validation		Independent Validation	
		Intended amount (ng)	Actual amount (ng)	Intended amount (ng)	Actual amount (ng)
KLF7^{ind}	72	-	-	200	200
TNFAIP8L3^{ind}	76	-	-	200	200
TEPP^{ind}	85	-	-	200	150
LRRC43	86	200	183	200	200

Target Region	Amplicon Size (bp)	Technical Validation		Independent Validation	
		Intended amount (ng)	Actual amount (ng)	Intended amount (ng)	Actual amount (ng)
ISM2	90	200	113	200	191
ZNF570	93	200	171	200	200
ZNF790	101	200	200	150	150
CACNA2D2	104	200	124	150	149
ELOVL5	119	200	45	150	66
TRANK1	122	200	176	150	150
ZNF699[ind]	125	-	-	200	200
TUBA3FP	137	200	200	200	200
SCOC	148	200	200	200	200
VANGL2[ind]	148	-	-	200	200
ZNF221[ind]	160	-	-	150	150
MGMT[ind]	164	-	-	150	150
ARL10	181	200	200	150	150
Upstream CA4[ind]	196	-	-	150	150

Following pooling, each patient sample (now consisting of amplified methylated and unmethylated target DNA) underwent clean up as per Section 2.2.3.2 Qiagen Min Elute PCR Purification Kit, recovering DNA down to 70bp in size. A 100 – 200ng aliquot of clean pooled DNA was taken per patient and stored at -20°C for later comparison against post-library preparation equivalent. Purified DNA was quantified using the broad range Qubit® kit, as outlined in Section 2.2.4 Qubit® fluorometric double-stranded DNA quantification. DNA was diluted to obtain 50µL at 20ng/µL for library preparation and stored at -20°C overnight.

5.2.4.2 MiSeq library preparation and quality control

For each patient, 1µg of pooled amplified target was used for MiSeq library preparation. Technical and independent validation cohorts were prepared separately. For unique indexing of 24 samples (adapter sequences given in Appendix 6: Illumina TruSeq Adapter Sequences), Illumina TruSeq DNA PCR-Free LT Sample Preparation Kit, Set A and Set B were required (Cat #FC-121-3001 and Cat #FC-121-3002 respectively). End-repair, A-tailing and adapter ligation as per manufacturer's protocol. Clean-up steps were altered to account for shorter amplicon size. Clean-up after end-repair used a 2:1

ratio of undiluted beads:DNA to ensure capture of <100bp fragments, with volumes adjusted accordingly. Double clean-up after adapter ligation used a 1:1 ratio of undiluted beads:DNA to capture adapter-ligated DNA fragments >200bp, and remove adapter dimers or incomplete ligation products.

A 60-100ng aliquot was taken from each post-library preparation sample (less DNA required due to longer fragment length) for comparison against the pre-library preparation equivalent. Pre- and post-library preparation DNA was visualized on 2.5% agarose SYBRsafe gel (as per protocol in Section 2.2.2 Agarose gel electrophoresis), to confirm the approximate 120-130bp increase in adapter ligated samples. The resultant adapter-ligated DNA was quantified using the Qubit® high sensitivity kit, as outlined in Section 2.2.4 Qubit® fluorometric double-stranded DNA quantification. Calculations were performed to convert ng/μL to nM for each sample, using an average library size of 238bp (technical cohort, 10 target regions) and 242bp (independent cohort, 18 target regions), using the following equation:

$$x\frac{ng}{\mu L}\times\frac{10^6\,\mu L}{1L}\times\frac{1nmol}{660ng}\times\frac{1}{N}=nM$$

where N = average fragment size (bp)
x = Qubit concentration (ng/μL)

30nM and 10nM single pooled samples were prepared using adapter ligated individual patient samples, adding the appropriate volume of each sample, diluting in Illumina resuspension buffer. 5μL of 30nM and 10nM pooled samples were taken for KAPA quantification (Section 5.2.4.3 Quantification of PCR-competent sequencing template). Pooled sample concentration was re-checked using the Qubit® high sensitivity kit, as outlined in Section 2.2.4 Qubit® fluorometric double-stranded DNA quantification.

5.2.4.3 Quantification of PCR-competent sequencing template

KAPA Biosystems Library Quantification Complete Kit, optimized for Roche LC480 (GeneWorks, Cat #KP-KK4854) was used for accurate quantification

of PCR-competent sequencing templates, crucial for reliable cluster amplification. Quantification was performed on single pooled 10nM and 30nM samples (1:100,000, 1:200,000, 1:400,000 and 1:800,000 dilutions), measured against a standard curve (20pm – 0.0002pm, 10-fold dilution series), as per the manufacturer's instruction. Reactions were performed in triplicate. Background subtracted amplification curves and Ct scores were reviewed, ensuring replicate data points differed by ≤ 0.2 cycles and all library dilutions fell within the dynamic range of the assay. A standard curve was generated, ensuring the following criteria were met: (i) Average ΔCt between standards in the range 3.1 – 3.6, (ii) Reaction efficiency 90 – 110%, (iii) $R^2 \geq 0.99$, (iv) NTC Ct detected at least 3 cycles later than average Ct for standard 6 (0.0002pm). KAPA Library Quantification Data Analysis Template was used to determine working concentration of single pooled 10nM and 30nM samples, including size adjustment calculations to account for difference between average fragment length of the library and KAPA standards (452bp).

5.2.4.4 MiSeq targeted sequencing run

Illumina MiSeq reagent Kit v2, 300 cycles (Cat #MS-102-2002) was used to prepare the library for sequencing with the MiSeq desktop sequencer (up to 325 cycles of sequencing, sufficient for up to 151-cycle paired-end reads plus two eight-cycle index reads). Library was diluted to 2nM working concentration (using KAPA quantification calculation) for denaturation with NaOH to a final concentration of 10pM DNA in 1mM NaOH. A 10nM PhiX (Illumina, PhiX Control v3, Cat #FC-110-3001) was prepared and denatured alongside the 10nM library (final concentration 10pM in 1mM NaOH). Illumina Experiment Manager (IEM) was used to create sample sheets and all kit components were prepared as per the manufacturer's instruction. A maintenance wash was performed on the MiSeq desktop sequencer prior to performing the run. Cluster density (K/mm) and cluster passing filter (%) results were recorded for each run.

5.2.5 MiSeq data analysis

MiSeq bioinformatic analysis was kindly performed by Dr Phuc Loi Luu at the Garvan Institute of Medical Research. Paired-end fastq files were obtained for each library and aligned to hg19 using GitHub bwa-meth (arXiv:1401.1129). Downstream analysis was performed using the 'ampliconAnalysis' function of the R v3.2.3 package aaRon. Data quality was assessed using on-target read counts and bisulfite conversion efficiency for all amplicons in all samples. Percent methylation at each CpG site was calculated and used in all subsequent analysis.

5.2.6 External differential methylation validation

Raw HM450 data and sample annotations were obtained from the University of Queensland study, Krause et al[49] (GSE72874; to be released 01/04/2016). The study included 250 samples from 154 patients, comprised of 125 EAC, 19 BE (8 from BE patients, 11 adjacent BE from EAC patients), 75 normal squamous esophagus and 31 control (10 GERD and 21 normal stomach samples); herein referred to as the 'External validation cohort' (Table 3-2). Raw HM450 data was obtained for peripheral blood samples (Blood_PT1-5) from Slieker et al[110] (PMID:23919675; GSE48472). Illumina HumanMethylation450 (HM450) raw data was pre-processed and background normalised with Bioconductor minfi package (PMID: 24478339), using the preprocessIllumina(…, bg.correct = TRUE, normalise = "controls", reference=1) command. Resultant M-values were used for differential methylation statistical analysis[105]. Target regions (as defined in Table 5-5) were elongated to include flanking regions either side, and were evaluated for average methylation status at each probe using the external validation cohort. Data was analyzed (i) using only BE samples from BE patients (n=8) (ii) using all BE samples (n=19); those from BE patients (n=8) and adjacent BE from EAC patients (n=11).

5.2.7 Chromatin state discovery and characterization

We used computational segmentation (ChromHMM v1.10[141], Core 15-state model) of the primary esophagus epigenome into chromatin states to assist

in characterization of regions for MiSeq amplicon sequencing. Segmentation was obtained from the Epigenome Roadmap project (http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html). Briefly, ChromHMM segmentation model was trained on 60 high-quality epigenomes spanning five chromatin marks H3K4me3, H3K4me1, H3K36me3, H3K27me3 and H3K9me3. The model was then applied to produce chromatin state assignment in 127 epigenomes including primary esophagus (E079; primary esophagus tissue from 34 year old Caucasian male).

5.2.8 Differential expression in selected target regions for validation

Differential expression was examined in genes corresponding to target regions selected for validation (regions as defined in Table 5-5). Transcript cluster ID, log fold change, t-statistic, p-value and transcript ID were recorded, with statistically significant differential expression defined as $p \leq 0.05$.

5.3 Results

5.3.1 Bisulfite-specific PCR for target region amplification

5.3.1.1 Target regions for assay design

Target regions for BSP assay design were based on biomarker filtering cut-offs and exclusion criteria applied to genome-wide methylation profiling data, as well as weighted priority ranking, based on baseline methylation, Δβ, and number of probes in the identified differentially methylated region (Section 5.2.1 Target region selection). Resultant target regions (identified by their RefSeq gene association) for BSP assay design are given in Table 5-4 below.

Table 5-4: Target regions for BSP assay design, identified by their RefSeq gene association. 26 regions associated with 25 genes (§: two distinct regions within KLF7) across three comparison classes were included with focus on the clinically relevant BE v HGD-EAC comparison for identification of intervention requiring disease.

BE v HGD-EAC		N v HGD-EAC	N v Any disease
ADAM22	TEPP	ELOVL5	MGMT
ARL10	TNFAIP8L3	PRDM2	RUNDC3B
C1orf51	TRANK1	ZNF790	SCOC
CACNA2D2	TUBA3FP		ZNF569
GRASP	Upstream CA4		ZNF570
ISM2	VANGL2		
KLF7[§]	ZNF221		
LRRC43	ZNF699		
MTERF			

MGMT, although not ranked highly in weighted priority lists, was detected in differential methylation analysis and was included for targeted amplicon sequencing as a result of considerable literature support: promoter hypermethylation of MGMT has been reported to have diagnostic and prognostic utility not only in esophageal adenocarcinoma and gastrointestinal cancers, but also in glioblastoma, lung and cervical cancers[91-93, 147-151]. Notable was the abundance of aberrantly methylated zinc finger loci across all comparison groups. Hypermethylation of zinc finger protein genes has long been associated with carcinogenesis in various forms of human cancer, however these particular loci and their role in esophageal adenocarcinoma development have yet to be elucidated.

5.3.1.2 Successful assays for BSP amplification

A total of 165 primers (forward and reverse) for assays over 26 target regions were designed and tested. Of these assays, a number were unable to be successfully optimized due to biased amplification of methylated/unmethylated input, non-specific product formation, amplification of wild-type (non-bisulfite converted) DNA input, and were discarded. Selected disease-associated differentially methylated target regions for BSP amplification and sequencing validation (18 of 26 target regions) are outlined in Table 5-5.

Table 5-5: Disease-associated, differentially methylated target regions for BSP amplification and sequencing validation, identified by RefSeq gene

association. Regions in grey were additional targets included for independent validation cohort (18 target regions; technical validation cohort evaluated in 10 target regions only). CpG site count refers to sites for methylation analysis only (CpG sites occurring in primer regions not analyzed).

	Locus	Chr	Start	End	Amplicon length (bp)	CpG sites
N v HGD-EAC						
ELOVL5	6p12.1	6	53,212,967	53,213,085	119	8
ZNF790	19q13.11	19	37,329,427	37,329,327	101	6
N v Any disease						
SCOC	4q31.1	4	141,294,938	141,294,791	148	10
ZNF570	19q13.11	19	37,960,451	37,960,359	93	5
MGMT	10q26.3	10	131,265,276	131,265,113	164	17
BE v HGD-EAC						
ARL10	5q35.2	5	175,792,604	175,792,424	181	19
CACNA2D2	3p21.31	3	50,541,229	50,541,332	104	8
ISM2	14q24.3	14	77,965,343	77,965,434	92	6
LRRC43	12q24.31	12	122,667,942	122,667,857	86	7
TRANK1	3p22.2	3	36,985,944	36,986,065	122	13
TUBA3FP	22q11.22	22	21,368,817	21,368,681	137	8
KLF7	2q33.3	2	208,031,073	208,031,144	72	8
TNFAIP8L3	15q21.3	15	51,385,011	51,385,936	76	4
TEPP	16q21	16	58,018,911	58,018,827	85	6
ZNF699	19p13.2	19	9,420,542	9,420,418	125	10
VANGL2	1q23.3	1	160,370,041	160,370,188	148	9
ZNF221	19q13.2	19	44,455,289	44,455,448	160	11
Upstream CA4	17q23.2	17	58,212,732	58,212,927	196	8

Primers for successfully optimized BSP assays averaged 25 ± 3bp (minimum 18bp, maximum 33bp), sequences given in Table 5-6.

Table 5-6: Primer sequences for successfully optimized BSP assays for target region amplicon sequencing. Sequences as ordered, specific for bisulfite converted DNA. Degenerate bases ‘Y’: pyrimidine (thymine or cytosine) and ‘R’: purine (adenine or guanine). ‘T’ indicates cytosine origin in wild-type DNA (prior to bisulfite conversion); ‘t’ thymine origin in wild-type DNA. Equivalently for ‘A’ and ‘a’ in reverse primers.

	Forward Primer	Reverse Primer
N v HGD-EAC		
ELOVL5	YgttgtTtggtttTaataaaaaTtg	AcctttccttttttAtcctAttctctc
ZNF790	ggatggggTtTaaatgTagg	AaacRAtAcactctAAAaaatAtaAtcc
N v Any disease		
SCOC	ggaggtaTtgaggatgTagTtTtag	ActAccRAAAtcaAttAAtccaa
ZNF570	gTaTYggagtTaTaTagatggTTtg	ccaccaAttctAttctAcaAAtc
MGMT	gTYgggTTtgaggTagtTtg	caaactaaAAcacaAaAcctcaAA
BE v HGD-EAC		
ARL10	agYgTTagTaTTaaggggTTTag	caAcRAcaAtAcattAtAAAActtAtaAtA
CACNA2D2	gTagTTYgaatTTtgTTTtaTtggTt	aAAaAcRAcaAAAtctAaAAtAtAAAAAc
ISM2	gTaTTtggaYgtTTTTTaTTTtTTtT	A cRAccaAatcaaaAaAaAcc
LRRC43	gTagatTTTaggTTTTagggaTTagTT	ccRAaactctAAAAAcctAAacc
TRANK1	TYgggaaTTTTtgaggat	AtccRAaAAAaccaAAAcac
TUBA3FP	gaaggTagggaggatgTTTaTaagg	cccRAcaacccaAtActAtaAAttA
KLF7	gggTtatataagggggTtgagg	acRaccctccctAtcccttA
TNFAIP8L3	gataTtgagTaaTaggaggaagTTaTTTaTagT	AaccRAtAAaAAtAAAcaAaAacc
TEPP	tTTTTaagggTagtTTtgTTtTTagatT	AAAccRctAaAatAaaActAtAAcacttA
ZNF699	gaaagaggaatTagaagtgtgTttTtg	AAAcRAAAcaAAAcaAAacaAAaaA
VANGL2	gtTTaggaagaTagTTTTaggagggaT	cctAcRccttcaAatAcccaA
ZNF221	gggTattTtgggaagtgtTTTaa	caActcaaaatatcacaAaataaaAAAtaac
Upstr. CA4	aaaTaTagaTtggaggagtggtaaataTtaa	tAtAccccaAaAaAcactAcaAAcc

5.3.1.3 Optimized BSP assay parameters

Bisulfite-specific PCR assays were optimized to amplify methylated and unmethylated DNA without bias prior to targeted sequencing using the MiSeq desktop sequencer. Optimized assay parameters are given in Table 5-7.

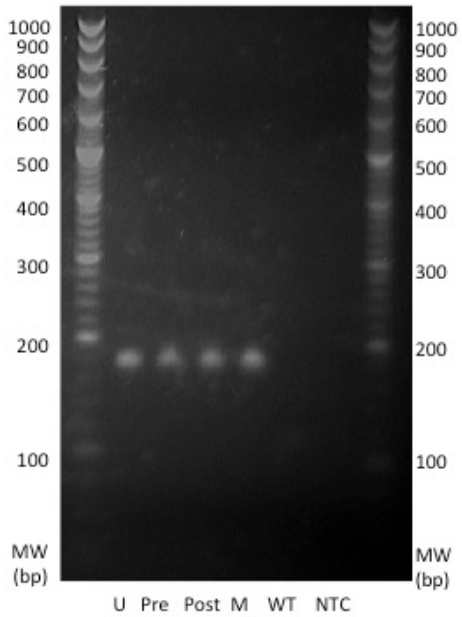
Table 5-7: Optimized BSP assay parameters for unbiased amplification of methylated and unmethylated DNA. An initial 5 cycles with longer anneal and extension times were always performed prior to 35-45 cycles of shorter anneal-extension. Tm: average melt temperature (°C) of 100% unmethylated (U) or methylated (M) DNA standard material, amplified using the stated parameters.

BSP Assay	Anneal (°C)	$MgCl_2$ (mM)	Amplification Cycles	Tm (°C) 100U	Tm (°C) 100M
ELOVL5	60	1.0	40	77.6 ± 0.2	81.4 ± 0.2
ZNF790	57	1.0	40	77.1 ± 0.4	80.1 ± 0.3
SCOC	57	1.0	40	78.9 ± 0.3	82.0 ± 0.2
ZNF570	56	1.0	40	72.2 ± 0.8	74.4 ± 0.8
ARL10	55	1.0	40	80.6 ± 0.4	86.8 ± 0.2
CACNA2D2	58	1.5	40	76.3 ± 0.3	81.1 ± 0.3
ISM2	57	1.5	35	76.9 ± 0.4	80.2 ± 0.1
LRRC43	56	1.0	40	75.1± 0.2	78.2 ± 0.3
TRANK1	55	1.0	40	82.1 ± 0.3	86.8 ± 0.2
TUBA3FP	61	1.0	40	77.2 ± 0.3	80.3 ± 0.4
KLF7	61	1.0	40	81.2 ± 0.4	87.2 ± 0.2
TNFAIP8L3	58	1.5	40	75.0 ± 0.3	77.4 ± 0.4
TEPP	60	1.5	40	74.5 ± 0.4	77.3 ± 0.3
ZNF699	54	1.0	45	74.8 ± 0.2	78.2 ± 0.3
VANGL2	60	1.0	40	79.2 ± 0.4	82.0 ± 0.2
ZNF221	61	1.0	40	80.8 ± 0.4	84.3 ± 0.2
MGMT	57	1.0	40	80.0 ± 0.5	84.6 ± 0.2
Upstr. CA4	62	1.0	45	77.7± 0.3	79.6 ± 0.3

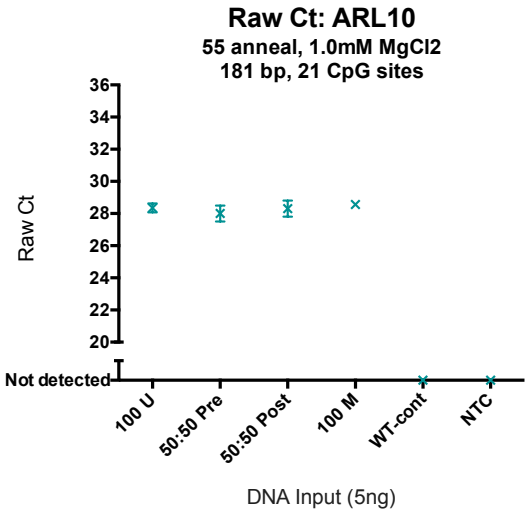
5.3.1.4 BSP assay quality control

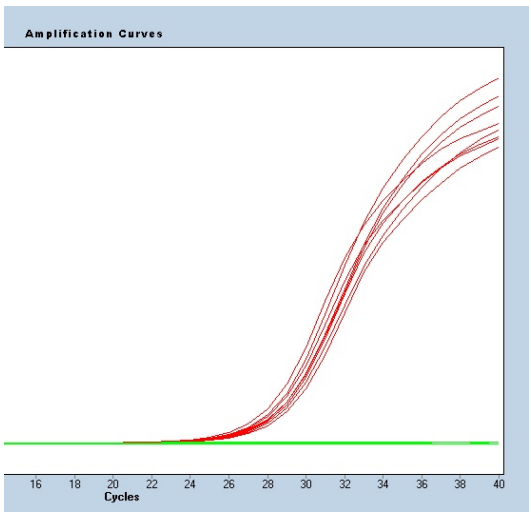
Optimization of bisulfite-specific PCR conditions results in amplification of a single product, of the expected size, without any amplification of non-specific products, wild type (non-bisulfite converted) control DNA or no template control. Characteristics of a successfully optimized bisulfite-specific PCR assay are outlined in Figure 5-1.

(A)

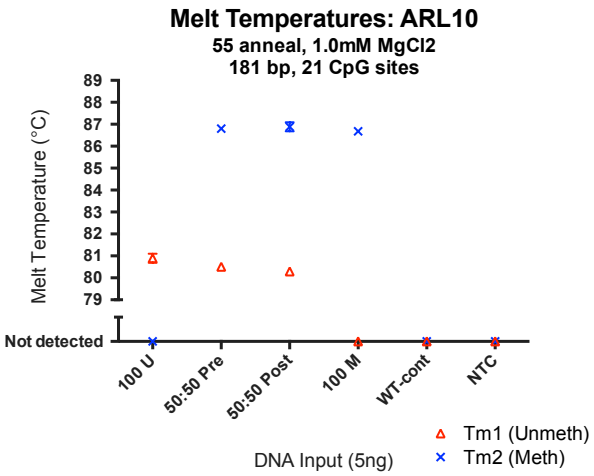

1000
900
800
700
600
500
400
300
200
100
MW (bp)
U Pre Post M WT NTC

(B)


Raw Ct: ARL10
55 anneal, 1.0mM MgCl2
181 bp, 21 CpG sites
Raw Ct
36
34
32
30
28
26
24
22
20
Not detected
100 U
50:50 Pre
50:50 Post
100 M
WT-cont
NTC
DNA Input (5ng)


Amplification Curves
16 18 20 22 24 26 28 30 32 34 36 38 40
Cycles

(C)


Melt Temperatures: ARL10
55 anneal, 1.0mM MgCl2
181 bp, 21 CpG sites
Melt Temperature (°C)
89
88
87
86
85
84
83
82
81
80
79
Not detected
100 U
50:50 Pre
50:50 Post
100 M
WT-cont
NTC
DNA Input (5ng)
Tm1 (Unmeth)
Tm2 (Meth)

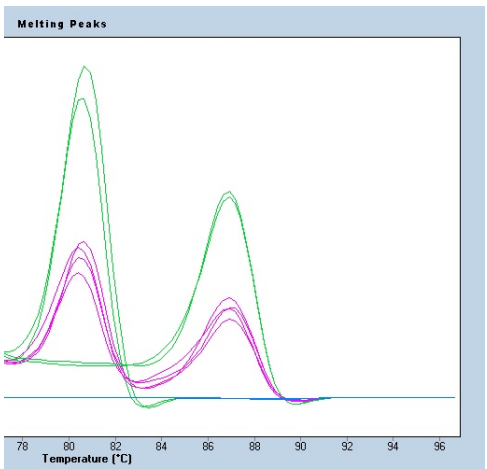

Melting Peaks
78 80 82 84 86 88 90 92 94 96
Temperature (°C)

Figure 5-1: Characteristics of a successfully optimized bisulfite specific PCR assay: 181bp region in Chr5: 175,792,604-175,792,424; HIG1 hypoxia inducible domain family, member 2A and ADP-ribosylation factor-like 10 (ARL10). (A) A clean, single product without evidence of non-specific product formation is detected for all methylated and unmethylated DNA inputs at the predicted molecular weight (MW): 181bp for ARL10. U: 100% unmethylated input, M: 100% methylated input, Pre: 50% methylated and 50% unmethylated input, mixed prior to bisulfite conversion, Post: 50% methylated and 50% unmethylated input, mixed following bisulfite conversion, clean-up and quantitation. No product is detected for non-bisulfite converted 'wild-type' DNA (WT) or no template control (NTC). Gel is 2.5% agarose, run at 100V constant for 90min. (B) Amplification of all DNA occurs at the same C_t (threshold cycle), ± < 1.0 C_t, irrespective of the methylation status of the DNA input. (C) Distinct melt temperatures (T_m) are obtained for methylated and unmethylated DNA input (melts in green). This region of ARL10 has 21 CpG sites, hence the large ΔT_m (~6°C). A 50:50 mixture of methylated and unmethylated DNA input (melts in pink) results in dual peaks at the same temperatures, and in proportion to those obtained for 100% methylated and unmethylated input. Results for (B) and (C) are the mean and standard deviation for duplicate reactions at the stated run conditions in Table 5-7.

5.3.2 MiSeq library preparation

5.3.2.1 Column versus bead-based clean up

MiSeq library preparation clean up steps (after end-repair and adapter ligation), were optimised to account for shorter amplicon sizes (resulting from design enabling direct transfer to analysis of cfc-DNA extracted from human blood plasma). Clean up post end-repair was tested using variable solid phase reversible immobilization (SPRI) bead ratios (1.6:1, 1.8:1, 2:1 undiluted beads:DNA, selected based on Figure 5-2) as well as column clean up (Qiagen Min Elute PCR Purification Kit, methods as per Section 2.2.3.2 Qiagen Min Elute PCR Purification Kit). Example amplicons were pooled into three samples for evaluation (80-90bp, 90-100bp, 180-190bp)

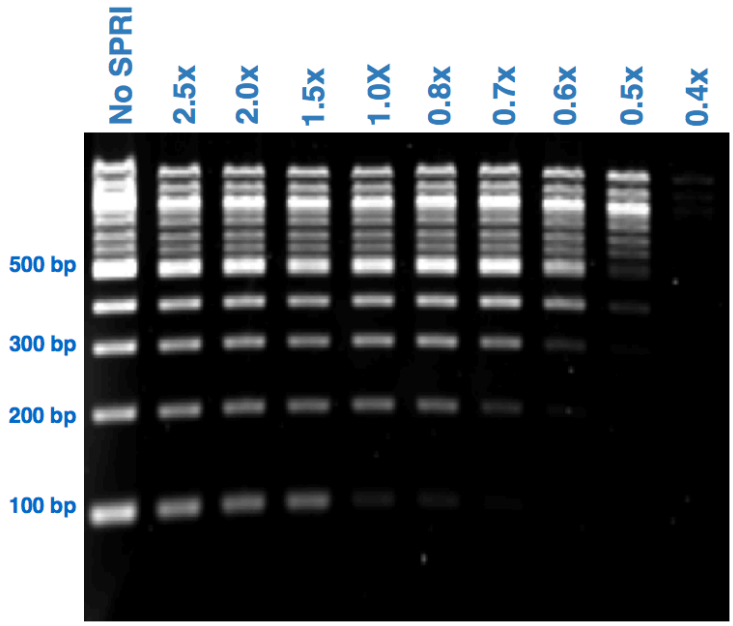


Figure 5-2: Retained DNA fragment sizes in relation to SPRI bead:DNA ratio. Image courtesy of CoreGenomics (http://core-genomics.blogspot.com.au/search?q=SPRI+beads).

Resultant DNA losses were calculated based on Qubit® broad range quantification (as per protocol Section 2.2.4 Qubit® fluorometric double-stranded DNA quantification), prior to and after clean-up. Clean-up success was evaluated via visualisation on 2.5% agarose (protocol Section 2.2.2 Agarose gel electrophoresis). Column clean up resulted in similar DNA loss across all input sizes (yields of 35.0, 29.9, 34.5% for 80-90bp, 90-100bp and 180-190bp pools respectively), whereas bead-based clean up yields were fragment size dependent, showing significant losses for the very short 80-90bp pool, irrespective of bead dilution, but very minimal loss for larger amplicons (Table 5-8). Agarose gel showed a clean, single band for all output (results not shown). Thus, in view of better performance for amplicons ≥ 90-100bp, bead-based clean up was implemented for MiSeq library preparation. A 2:1 ratio was used to minimize DNA loss, accepting that primer-dimer or self-dimer remaining at this point may not be removed (column based clean up prior to library preparation removes primer-dimer, self-dimer and excess primer remaining from amplification).

Table 5-8: Percentage yield based on DNA fragment size input and SPRI bead:DNA ratio used for clean-up. Insufficient very small fragment (80-90bp) DNA existed for testing of all bead ratios as well as column clean up, thus 2:1 ratio was omitted.

DNA fragment range (bp)	SPRI bead:DNA	Yield (%)
80 - 90	1.6:1	19.35
	1.8:1	24.75
90-100	1.6:1	32.90
	1.8:1	37.40
	2:1	44.80
180-190	1.6:1	91.50
	1.8:1	91.50
	2:1	102.75

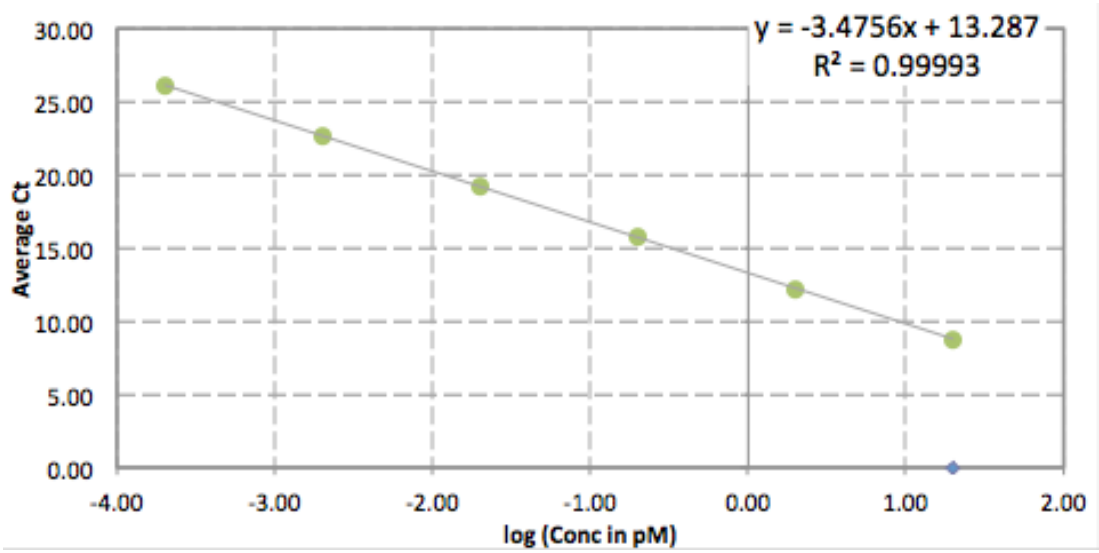
5.3.2.2 KAPA quantification

Accurate quantification of PCR-competent sequencing templates is crucial for reliable cluster amplification. For this, KAPA Biosystems Library Quantification complete kit, optimized for Roche LC480 was used to quantify 10nM and 30nM libraries for both technical and independent cohorts. Figure 5-3 shows KAPA quantification results for 10nM and 30nM libraries prepared for targeted sequencing of the technical validation cohort.

(A)

Well	Std #	Conc (pM)	Ct	Av Ct	Difference	Delta Ct
A1	1	20	8.73	8.76	-0.03	-
A2	1	20	8.77		0.01	
A3	1	20	8.77		0.01	3.43
B1	2	2	12.15	12.18	-0.03	
B2	2	2	12.20		0.02	
B3	2	2	12.20		0.02	3.60
C1	3	0.2	15.76	15.78	-0.02	
C2	3	0.2	15.83		0.05	
C3	3	0.2	15.76		-0.02	3.42
D1	4	0.02	19.22	19.21	0.01	
D2	4	0.02	19.22		0.01	
D3	4	0.02	19.18		-0.03	3.51
E1	5	0.002	22.72	22.71	0.01	
E2	5	0.002	22.72		0.01	
E3	5	0.002	22.70		-0.01	3.37
F1	6	0.0002	25.94	26.08	-0.14	
F2	6	0.0002	26.08		0.00	
F3	6	0.0002	26.23		0.15	3.92
H1	NTC	-	30.00	30.00		
H2	NTC	-	30.00			
H3	NTC	-	30.00			

(B)



(C)

Sample name	Dilution	Cq	Average fragment length (bp)	Average Cq	Difference	Delta Cq	log (conc)	Average conc (pM)	Size-adjusted conc (pM)	Undiluted library conc (pM)	Undiluted library conc (ng/µL)	% Deviation	Working conc (nM)	Working conc (ng/µL)
10nM library	100000	16.68	260	16.70	-0.02	-	-0.98	0.104	0.181	18,081	2.905	-	18.37	2.95
		16.75			0.05									
		16.68			-0.02									
	200000	17.76		17.76	0.00	1.05	-1.29	0.052	0.090	17,996	2.891	0.47%		
		17.77			0.01									
		17.74			-0.02									
	400000	18.79		18.79	0.00	1.03	-1.58	0.026	0.045	18,151	2.916	-0.39%		
		18.80			0.01									
		18.78			-0.01									
	800000	19.75		19.75	0.00	0.96	-1.86	0.0138	0.024	19,261	3.094	-6.53%		
		19.78			0.03									
		19.71			-0.04									
30nM library	100000	15.28	260	15.21	0.07	-	-0.55	0.280	0.486	48,627	7.812	-	47.88	7.69
		15.08			-0.13									
		15.27			0.06									
	200000	16.30		16.22	0.08	1.01	-0.84	0.143	0.249	49,809	8.002	-2.43%		
		16.31			0.09									
		16.05			-0.17									
	400000	17.32		17.35	-0.03	1.13	-1.17	0.068	0.118	47,017	7.554	3.31%		
		17.31			-0.04									
		17.43			0.08									
	800000	18.47		18.43	0.04	1.08	-1.48	0.0331	0.058	46,080	7.403	5.24%		
		18.34			-0.09									
		18.48			0.05									

Figure 5-3: KAPA quantification results for 10nM and 30nM libraries prepared for targeted sequencing of the technical validation cohort. (A) All library dilutions fell within assay dynamic range, with standard curve replicate data points differing by less than the required ≤ 0.2 cycles (average (abs) 0.033 ± 0.042 cycles) and average ΔCt 3.47 ± 0.09 cycles (within the required 3.1 – 3.6). (B) Generated standard curve reaction efficiency was 94% (required 90-110%), R^2 = 0.99993 (required R^2 ≥ 0.99), NTC detected 3.92 cycles later than 0.0002pm standard (required > 3.0 cycles later). (C) Following size adjustment calculations for difference between average library fragment length and KAPA standards, PCR competent sequencing template was determined to be at 18.37nM (2.95ng/µL) and 47.88nM (7.69ng/µL) for original 10 and 30nM libraries respectively.

5.3.2.3 MiSeq library quality control

Pooled amplified target DNA for each patient sample pre- and post-library preparation was visualized on 2.5% agarose gel to confirm the approximate 120-130bp increase in adapter ligated samples. All gels confirmed the expected increase in both technical and independent validation cohorts, however clear bands were difficult to visualize for smaller amplicons. Minimum on-target reads for all patient samples across all targets (including the smallest amplicons) were above the required minimum for robust methylation determination (Table 5-9); hence visualization difficulties are likely due to gel resolution, rather than yield loss during library preparation.

An example quality control gel (technical cohort, 10 target regions, 12 patient samples, pre- and post-library preparation) is shown in Figure 5-4 below.

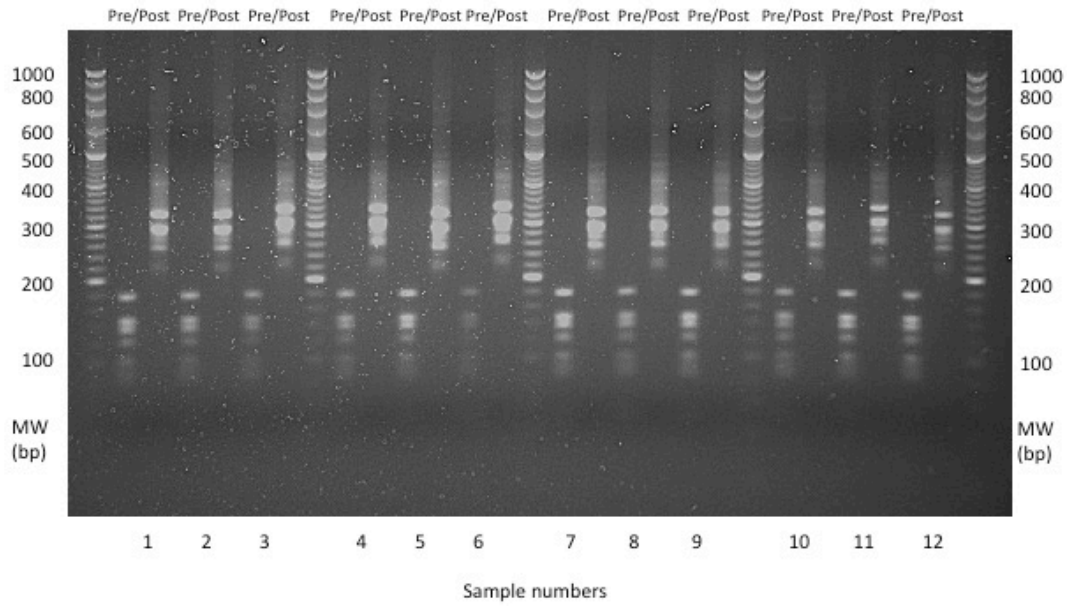


Figure 5-4: Agarose gel size verification of pooled amplicons pre- and post-library preparation. Twelve patient samples (N, BE) from the technical validation cohort, analyzing 10 target regions are shown. Pre-library preparation DNA ranges from 86 to 181bp (86, 92, 93, 101, 104, 119, 122, 138, 148, 181bp). Following adapter ligation there will be an approximate 120-130bp increase in fragment length. Due to size similarity some bands are unable to be individually distinguished.

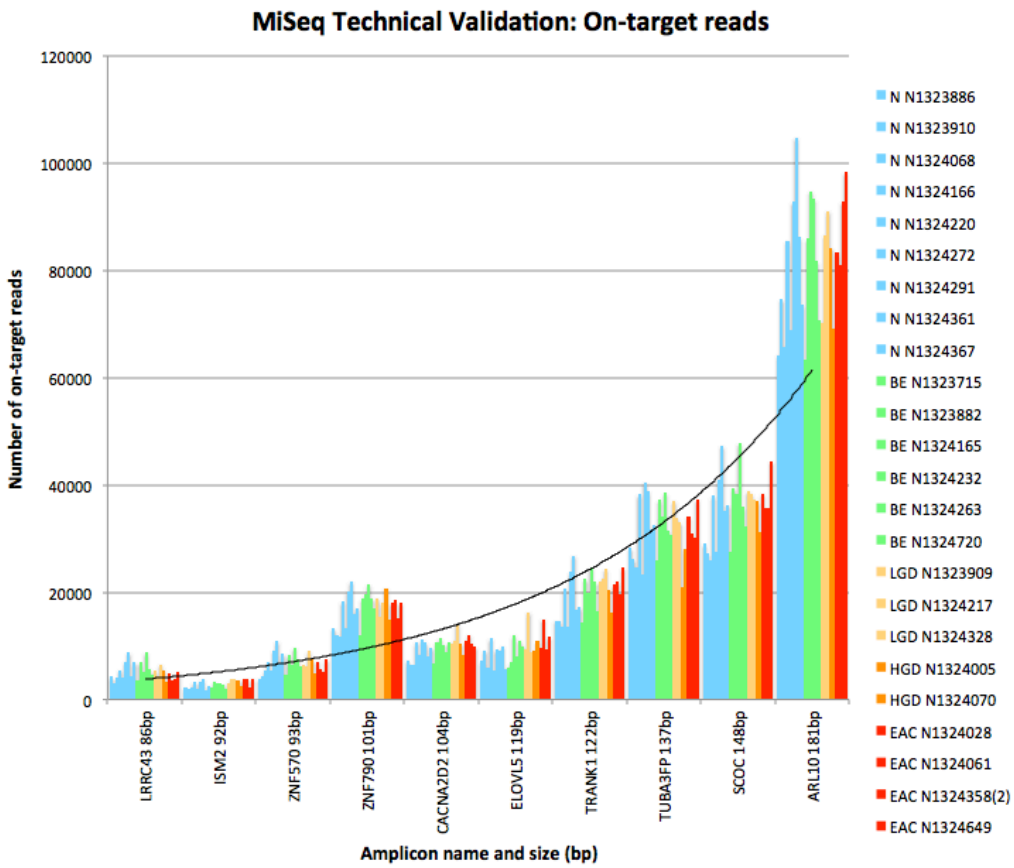
5.3.3 Robust methylation determination

Cluster density (K/mm) and cluster passing filter (%) of 592K/mm, 97.4% and 673K/mm, 96.8% were obtained for technical and independent validation respectively. Cluster passing filter should be >90%, ideally ≥ 95% for high quality data. Cluster densities of ~800K/mm are expected for libraries with pre-adapter ligation lengths > 200bp, however clustering is much less efficient for amplicons 100-150bp, as we have here (average pre-adapter ligation length 118 ± 30bp and 123 ± 37bp for technical and independent validation libraries). An average of 362,081 ± 55,280 total reads per patient sample (all target regions) was attained for technical validation; on-target reads comprising 61.04 ± 0.85%. A slight improvement in data quality was

attained for independent validation, averaging of 425,206 ± 72,985 total reads per patient sample; on-target reads comprising 65.75 ± 1.48%.

For robust methylation determination, >300 on-target reads are necessary. The number of on-target reads attained for each amplicon within each sample is impacted not only by initial amount of amplified target input into library preparation (150-200ng/assay maximum, refer to Table 5-3), but also by amplicon size (proportionally more loss of smaller amplicons occurs during clean up). This is apparent in Figure 5-5, showing on-target reads attained for each sample, grouped by amplicon. Pooling for independent validation was done in two batches (small (72-122bp) and large (125-196bp) amplicons), with input amounts (150/200ng maximum) varied to compensate for DNA loss due to amplicon size (Table 5-9).

(A)



(B)

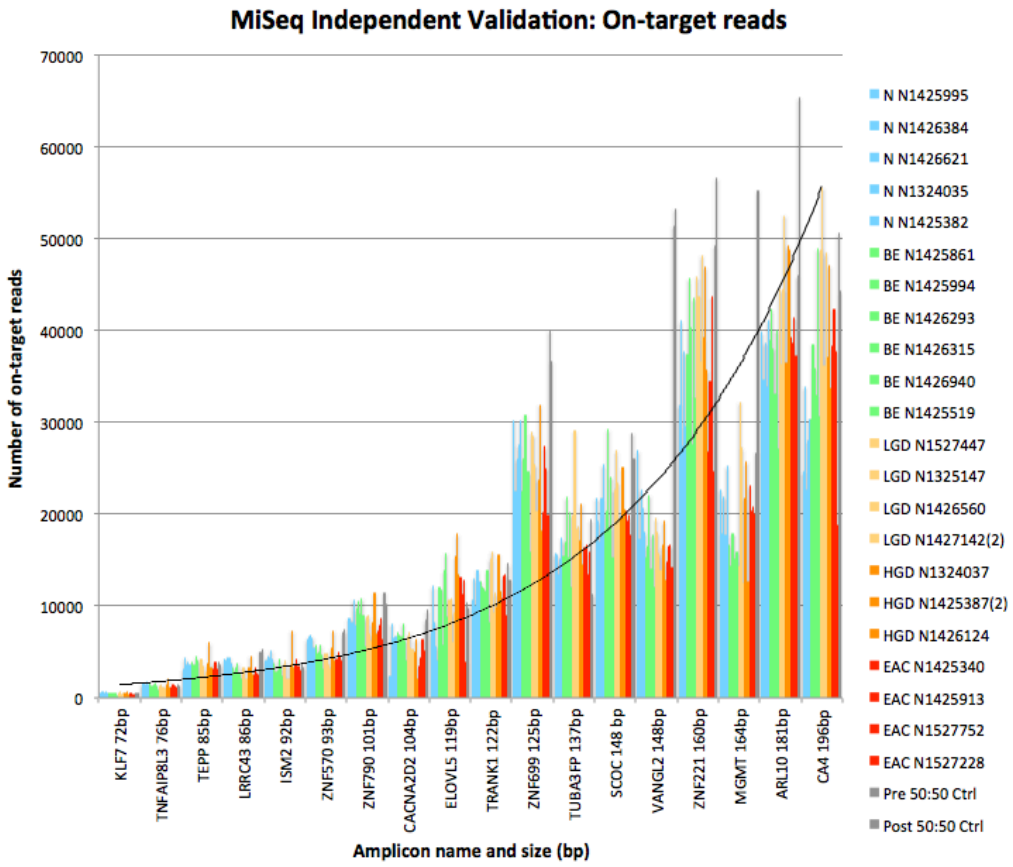


Figure 5-5: Number of on-target reads attained per patient sample for technical (A) and independent (B) validation. Results grouped by amplicon. On-target read count is impacted by amplicon length and amount of amplified DNA target input into library preparation.

For technical validation (10 target regions), a worst case of 1,732 on-target reads (amplification of ISM2 (92bp)), was attained for sample N1324361 (normal esophagus). Minimum on-target reads averaged 15,639; minimum 1,732; maximum 63,443 for this cohort. 1,732 on-target reads was well in excess of the minimum required for robust methylation determination; hence analysis of additional target regions was possible without impacting data quality. For this reason, a further 8 target regions were included (total 18) for independent validation. A worst case of 312 on-target reads (amplification of KLF7 (72bp)), was attained for sample N1425519 (non-dysplastic BE). Minimum on-target reads averaged 9,365; minimum 312; maximum 27,099.

Despite a significant increase in the number of target regions for analysis in the independent validation cohort (total library DNA remains unchanged), there were still sufficient on-target reads in every sample, for every target to enable robust methylation determination. Table 5-9 gives minimum on-target reads attained for both technical (10 target regions) and independent (18 target regions) validation.

Table 5-9: MiSeq targeted sequencing minimum on-target reads per patient sample for technical (10 target regions) and independent (18 target regions) validation. Regions in grey for independent validation only. 24 samples each cohort. Despite the increased number of target regions analyzed for independent validation, minimum on-target reads were still above the 300 required for robust methylation determination.

		Technical validation		**Independent validation**	
Target region	**Amplicon Size (bp)**	**DNA input to library prep (ng)**	**Minimum on-target reads**	**DNA input to library prep (ng)**	**Minimum on-target reads**
LRRC43	86	183	3,000	200	2,011
ISM2	92	113	1,732	191	2,111
ZNF570	93	171	3,859	200	3,097
ZNF790	101	200	11,840	150	5,677
CACNA2D2	104	124	6,467	150	2,150
ELOVL5	119	45	5,537	66	3,980
TRANK1	122	176	13,453	150	8,265
TUBA3FP	137	200	21,036	200	11,302
SCOC	148	200	26,022	200	15,251
ARL10	181	200	63,443	150	27,099
KLF7	72			200	312
TNFAIP8L3	76			200	924
TEPP	85			150	2,275
ZNF699	125			200	15,946
VANGL2	148			200	12,035
ZNF221	160			150	24,734
MGMT	164			150	12,597
Upstream CA4	196			150	18,814

5.3.4 Targeted amplicon sequencing validation

Average methylation at each CpG site within each target region was plotted for all patient samples analyzed (Table 5-10). Technical and independent validation cohorts were evaluated separately for cross-cohort, as well as genome-wide methylation profiling validation.

Table 5-10: Sample classification of technical and independent validation cohorts for MiSeq targeted amplicon sequencing. The combined HGD-EAC group was separated to determine if target region hypermethylation increases further with progression to invasive adenocarcinoma.

	Technical Validation Cohort	**Independent Validation Cohort**	**Total**
Normal	9	5	14
Non-dysplastic BE	6	6	12
LGD	3	4	7
HGD	2	3	5
EAC	4	4	8
Methylation controls	0	2	2
Totals:	**24**	**24**	**48**

Overall success or failure of the validation is summarized in Table 5-11 below. The majority of target regions validated were hypermethylated in HGD and EAC with respect to BE (BE v HGD-EAC, 13 regions), two regions hypermethylated in HGD and EAC with respect to normal esophageal mucosa (N v HGD-EAC) and a further three regions hypermethylated in BE, HGD and EAC with respect to normal esophageal mucosa (N v Any disease).

Table 5-11: Overview of success or failure of technical and/or independent validation of target regions examined by targeted sequencing. 18 target regions were validated, 10 in the technical validation, 18 in the independent validation (8 independent validation only targets in grey). A failed result is indicated if any data collected did not support original hypotheses from whole genome methylation profiling.

	Result	Reason
N v HGD-EAC		
ELOVL5	Success	All validation results supported whole genome methylation profiling
ZNF790	Success	All validation results supported whole genome methylation profiling
N v Any disease		
SCOC	Variable	Amplification bias, methylation above expected baseline in N samples
ZNF570	Failure	Methylation above expected baseline (N samples, independent cohort only)
MGMT	Success	All validation results supported whole genome methylation profiling
BE v HGD-EAC		
ARL10	Success	All validation results supported whole genome methylation profiling
CACNA2D2	Success	All validation results supported whole genome methylation profiling
ISM2	Failure	Independent cohort failed to validate technical results
LRRC43	Success	All validation results supported whole genome methylation profiling
TRANK1	Variable	Amplification bias, methylation above expected baseline in N samples
TUBA3FP	Success	All validation results supported whole genome methylation profiling
KLF7	Success	All validation results supported whole genome methylation profiling
TNFAIP8L3	Variable	Amplification bias, methylation above expected baseline in N samples
TEPP	Success	All validation results supported whole genome methylation profiling
ZNF699	Success	All validation results supported whole genome methylation profiling
VANGL2	Success	All validation results supported whole genome methylation profiling
ZNF221	Success	All validation results supported whole genome methylation profiling
Upstream CA4	Success	All validation results supported whole genome methylation profiling

Due to large amounts of data collected, example target regions will be examined in detail in this and the following sections, as outlined below:

- Examples of successful internal targeted amplicon sequencing validation (ARL10, TUBA3FP) in Section 5.3.4.1
- A brief discussion regarding failed and variable success validation targets (Examples: ISM2, TRANK1, TNFAIP8L3), in Section 5.3.4.2
- Further in-depth examination, including external validation in large cohort (Examples: KLF7, LRRC43, VANGL2), in Section 5.3.5
- Promoter methylation of MGMT (both internally and externally validated) in Section 5.3.5.1

With the exception of MGMT, no further in depth analysis was performed of the five targets in the N v HGD-EAC and N v Any disease comparisons, due to inability to differentiate non-dysplastic BE from dysplastic disease and EAC (not clinically relevant). Promoter methylation of DNA repair gene MGMT, due to its prevalence in literature and reported diagnostic and

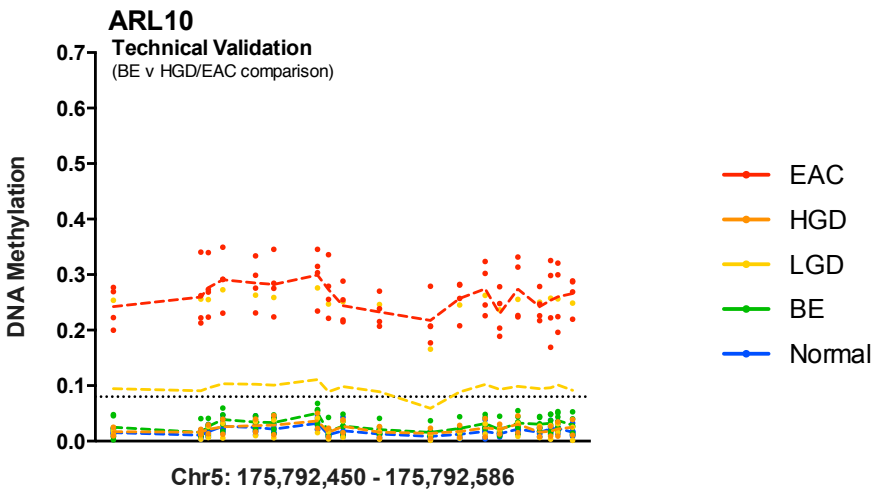
prognostic utility in esophageal adenocarcinoma and other cancers[93, 152, 153], is discussed further in Section 5.3.5.1.

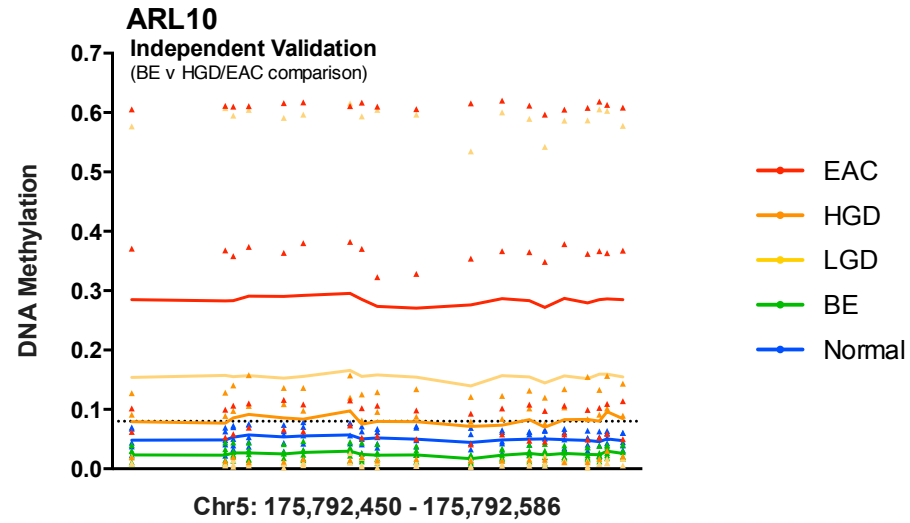
5.3.4.1 Successful internal sequencing validation

This section comprises an examination of targets with successful validation by MiSeq targeted amplicon sequencing, using examples ARL10 and TUBA3FP (both validated using technical and independent cohorts).

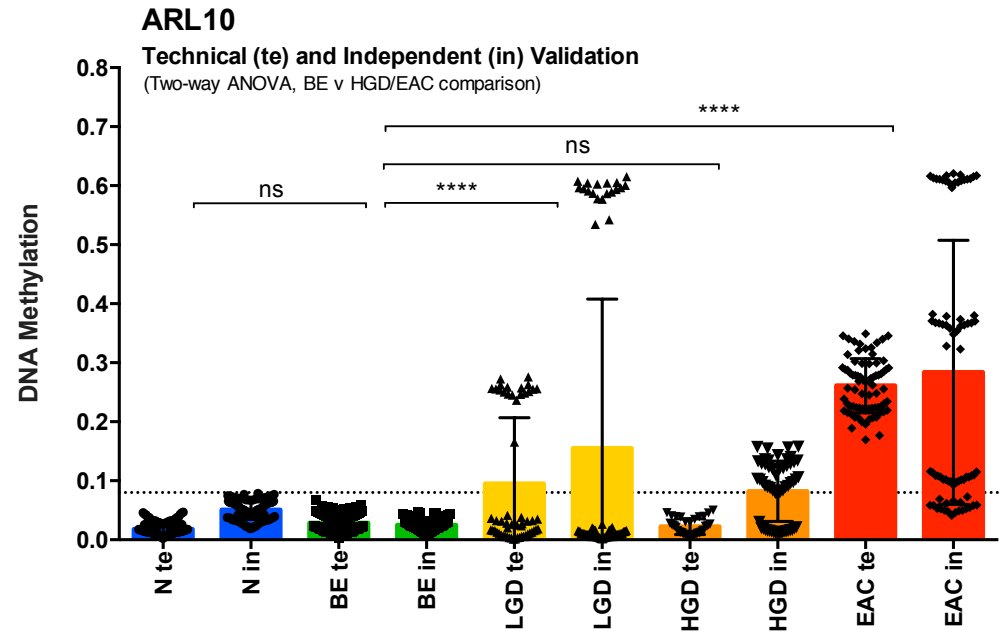
To date, there is not much known about ADP-ribosylation factor-like 10 (ARL10) and its role in carcinogenesis. In 2004, Louro et al[154], examined transcript levels of RASL11A and ARL9 (RASL11A and ARL9/ARL10 subfamily code for small GTPases, highly conserved among eukaryotes) in normal and prostate tumor samples; finding RAS11A to be significantly down-regulated in prostate tumor compared to normal prostate tissue, but no difference in ARL9 expression (in the same tissue samples). No methylation analysis was performed. Small guanosine-5'-triphosphate (GTP)-binding protein regulators (such as ARL10) act as molecular switches, regulating function of other proteins to control signaling pathways[155]. In view of the importance of small GTPases in the cell cycle, their interaction with a range of signaling proteins and the apparent adenocarcinoma-associated promoter methylation of ARL10; further studies into the functional mechanism of ARL10 are warranted.

(A)

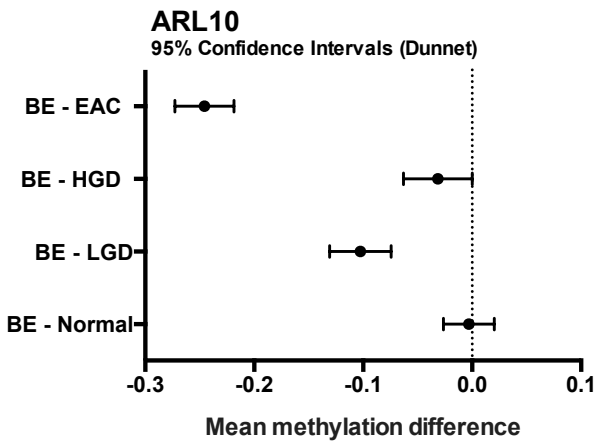



ARL10
Independent Validation
(BE v HGD/EAC comparison)
DNA Methylation
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
EAC
HGD
LGD
BE
Normal
Chr5: 175,792,450 - 175,792,586

(B)


ARL10
Technical (te) and Independent (in) Validation
(Two-way ANOVA, BE v HGD/EAC comparison)
DNA Methylation
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
ns
****
ns
****
N te
N in
BE te
BE in
LGD te
LGD in
HGD te
HGD in
EAC te
EAC in

(C)


ARL10
95% Confidence Intervals (Dunnet)
BE - EAC
BE - HGD
BE - LGD
BE - Normal
-0.3
-0.2
-0.1
0.0
0.1
Mean methylation difference

(D)

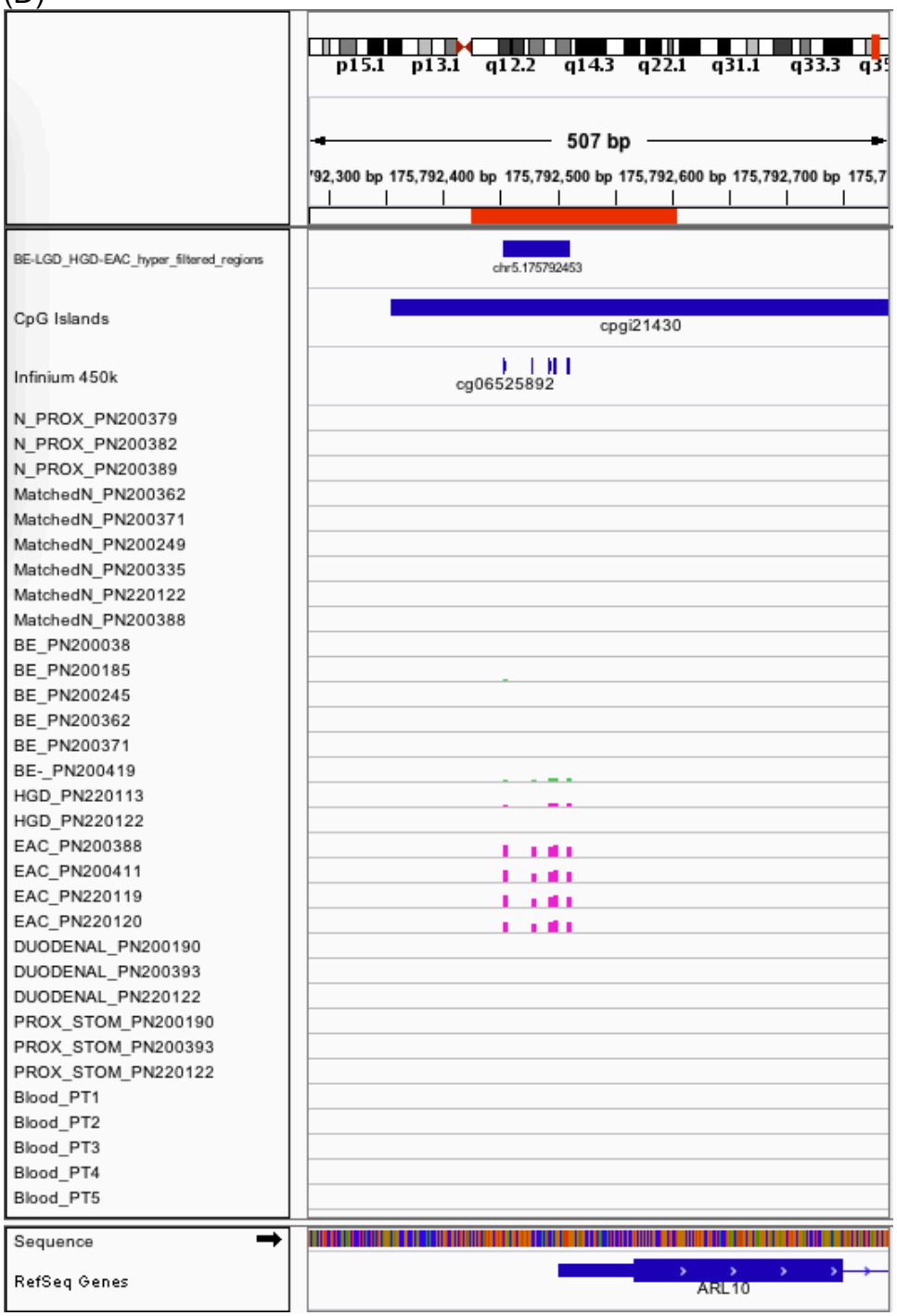
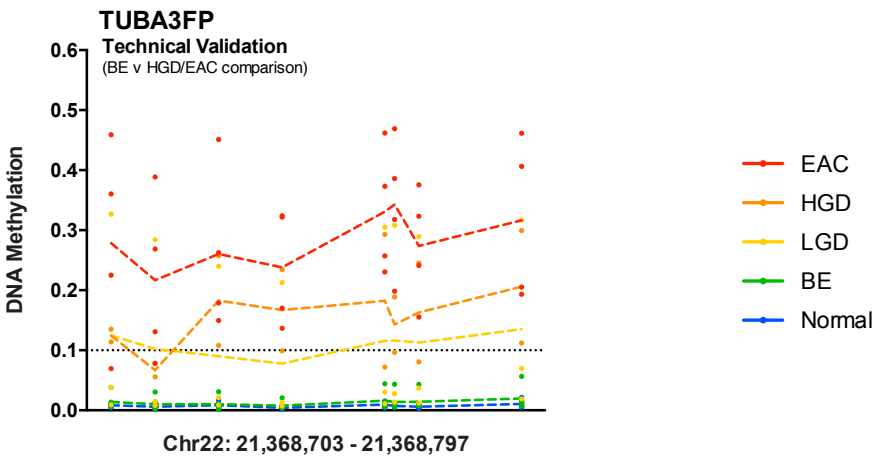


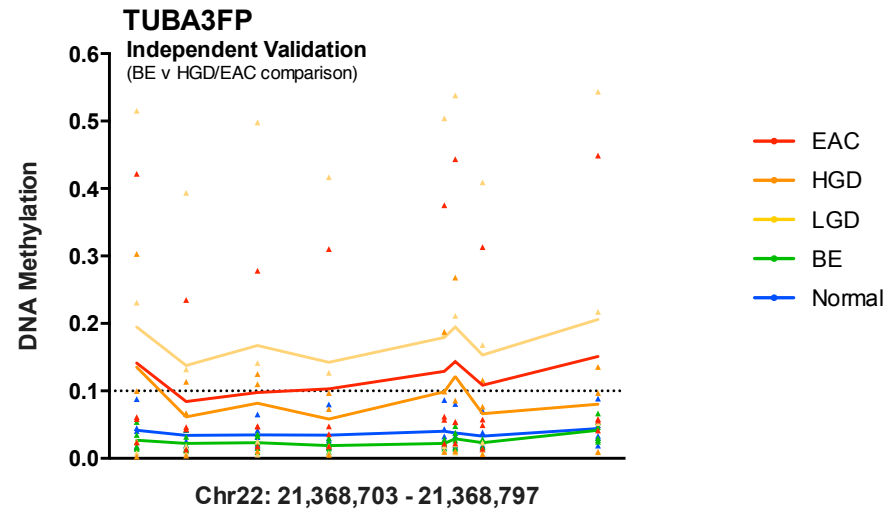
Figure 5-6: Targeted amplicon sequencing of disease-associated differential methylation occurring in the promoter region of ADP-ribosylation factor-like 10 (ARL10). (A) Targeted amplicon sequencing region (nineteen CpG sites, Chr5: 175,792,604 – 175,792,424) showing results for technical and independent validation cohorts separately. (B) Hypermethylation of a subset of dysplastic disease and EAC observed for the technical cohort was validated by independent cohort results. The LGD patient (independent

cohort) with observed ~60% methylation across all CpG sites within the target region had no further visits on clinical follow up, therefore predicted disease progression was unable to be confirmed. (C) Mean methylation difference (technical and independent sequencing data) from non-dysplastic BE was the most significant for EAC. (D) Technical and independent sequencing validate the observed aberrant methylation detected using genome-wide methylation profiling; genome browser showing HM450 data for a subset of the Training Cohort with >90% tissue sample homogeneity; ARL10 amplicon shown in red. Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

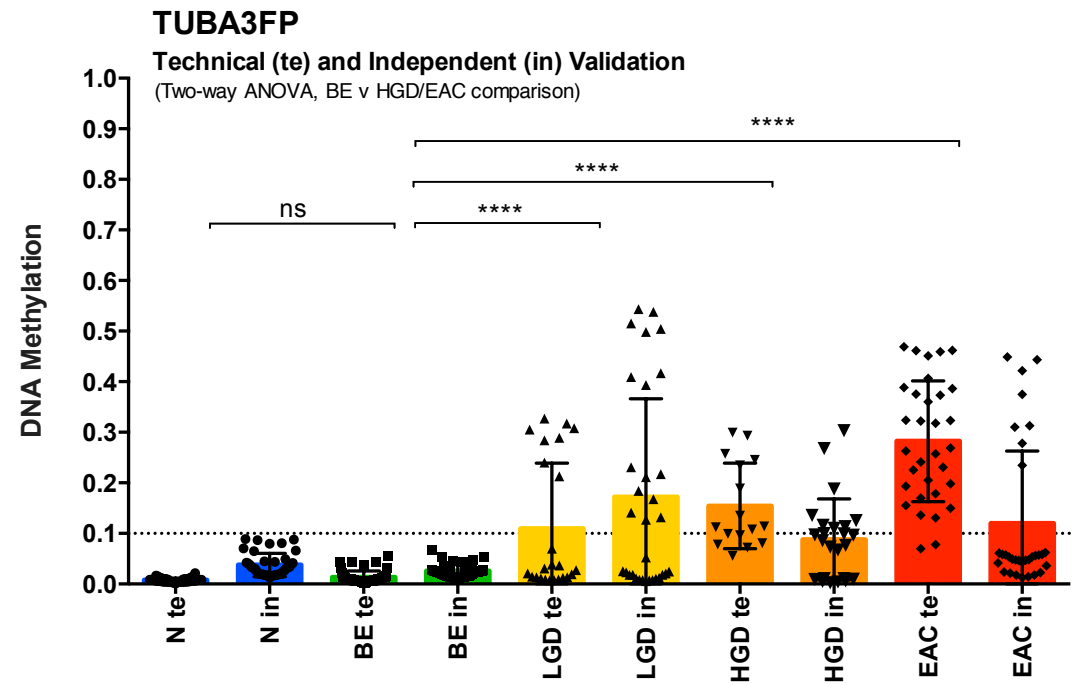
Aberrant promoter hypermethylation (our target for sequencing validation positioned within the flanking active transcription start site, Figure 5.7(D)) of Tubulin, alpha 3f, pseudogene (TUBA3FP) in EAC is a novel finding. Pseudogenes, although not protein-coding, may still be functional or have a regulatory role. There is not much known regarding pathways, gene ontology, biological processes or associated disorders in which TUBA3FP is involved, however there is some evidence for the use of TUBA3FP for determination of cancer tissue origin, scoring genes from exome-based analyses for level of mutational burden, presented by Robasky et al at the 2013 ASHG (American Society of Human Genetics) conference (http://www.ashg.org/2013meeting/abstracts/fulltext/f130123239.htm).

(A)

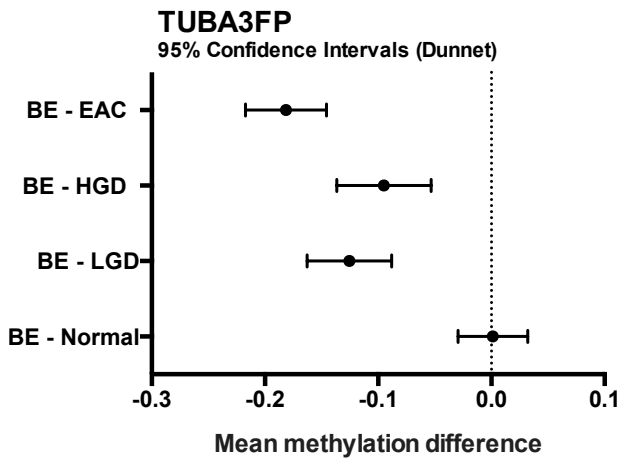



TUBA3FP
Independent Validation
(BE v HGD/EAC comparison)
DNA Methylation
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Chr22: 21,368,703 - 21,368,797
EAC
HGD
LGD
BE
Normal

(B)


TUBA3FP
Technical (te) and Independent (in) Validation
(Two-way ANOVA, BE v HGD/EAC comparison)
DNA Methylation
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
ns
****
****
****
N te
N in
BE te
BE in
LGD te
LGD in
HGD te
HGD in
EAC te
EAC in

(C)


TUBA3FP
95% Confidence Intervals (Dunnet)
BE - EAC
BE - HGD
BE - LGD
BE - Normal
-0.3
-0.2
-0.1
0.0
0.1
Mean methylation difference

(D)

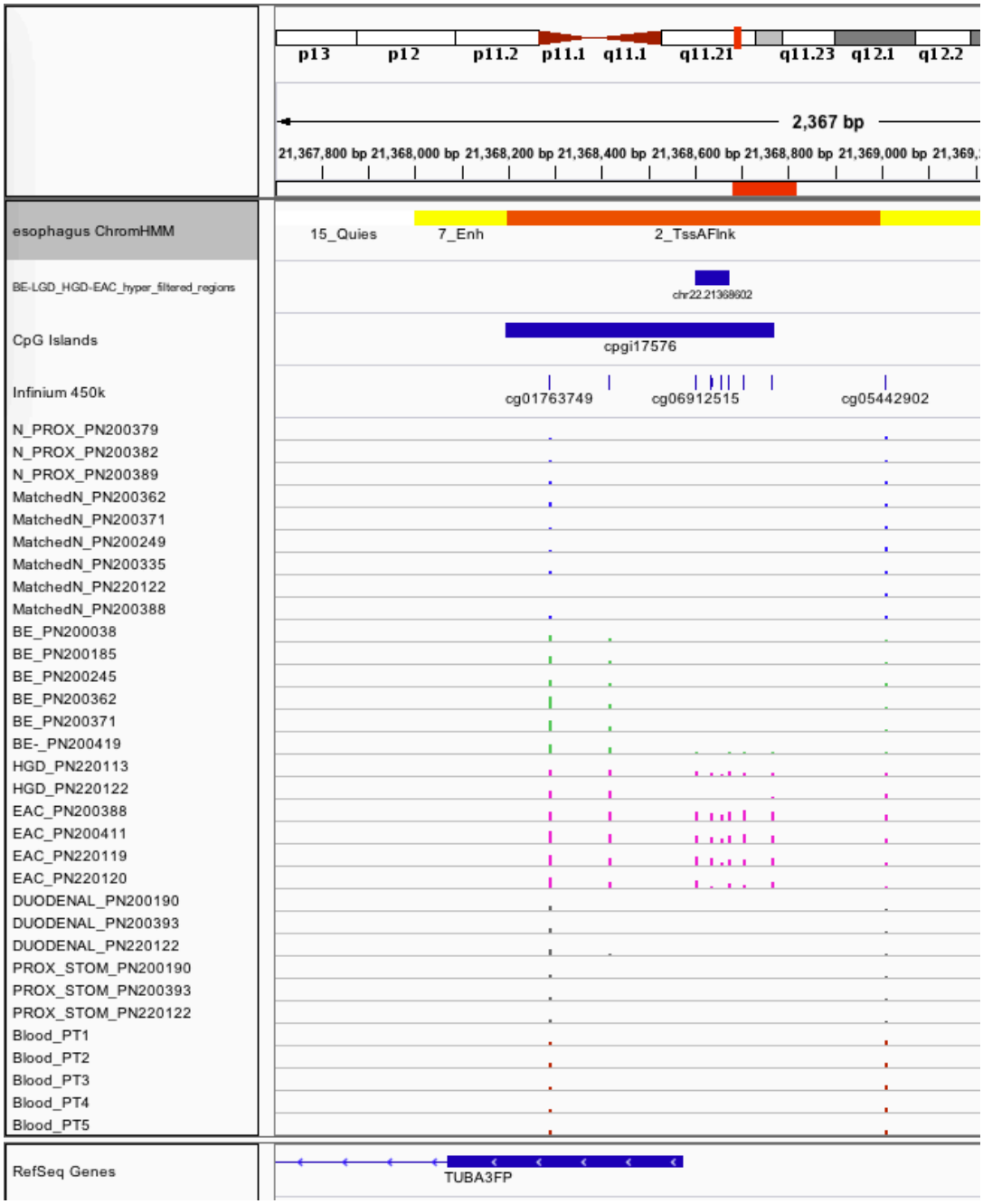


Figure 5-7: Targeted amplicon sequencing of disease-associated differential methylation occurring in the promoter region of Tubulin, alpha 3f, pseudogene (TUBA3FP). (A) Targeted amplicon sequencing region (eight CpG sites, Chr22: 21,368,817 – 21,368,681) showing results for technical and independent validation cohorts separately. Methylation level for dysplastic and EAC samples were elevated in comparison to non-dysplastic BE and N in the independent cohort, but not to the same extent (Δβ) as the technical cohort. (B) Hypermethylation of a subset of dysplastic disease and

EAC observed for the technical cohort was validated by independent cohort results. Hypermethylation of the entire EAC technical cohort was not reproducible in the independent cohort. (C) Mean methylation difference (technical and independent sequencing data) from non-dysplastic BE was the statistically significant for not only EAC, but also any form of dysplastic Barrett's. (D) Technical and independent sequencing validate the observed aberrant methylation detected using genome-wide methylation profiling; genome browser showing HM450 data for a subset of the Training Cohort with >90% tissue sample homogeneity; TUBA3FP amplicon shown in red. Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

5.3.4.2 Failed and variable success validation targets

This section comprises an examination of targets with failed or variable success validation, using examples of ISM2, TRANK1 and TNFAIP8L3 (excerpt from Table 5-11 below).

	Result	Reason
ISM2	Failure	Independent cohort failed to validate technical results
TRANK1	Variable	Amplification bias, methylation above expected baseline in N samples
TNFAIP8L3	Variable	Amplification bias, methylation above expected baseline in N samples

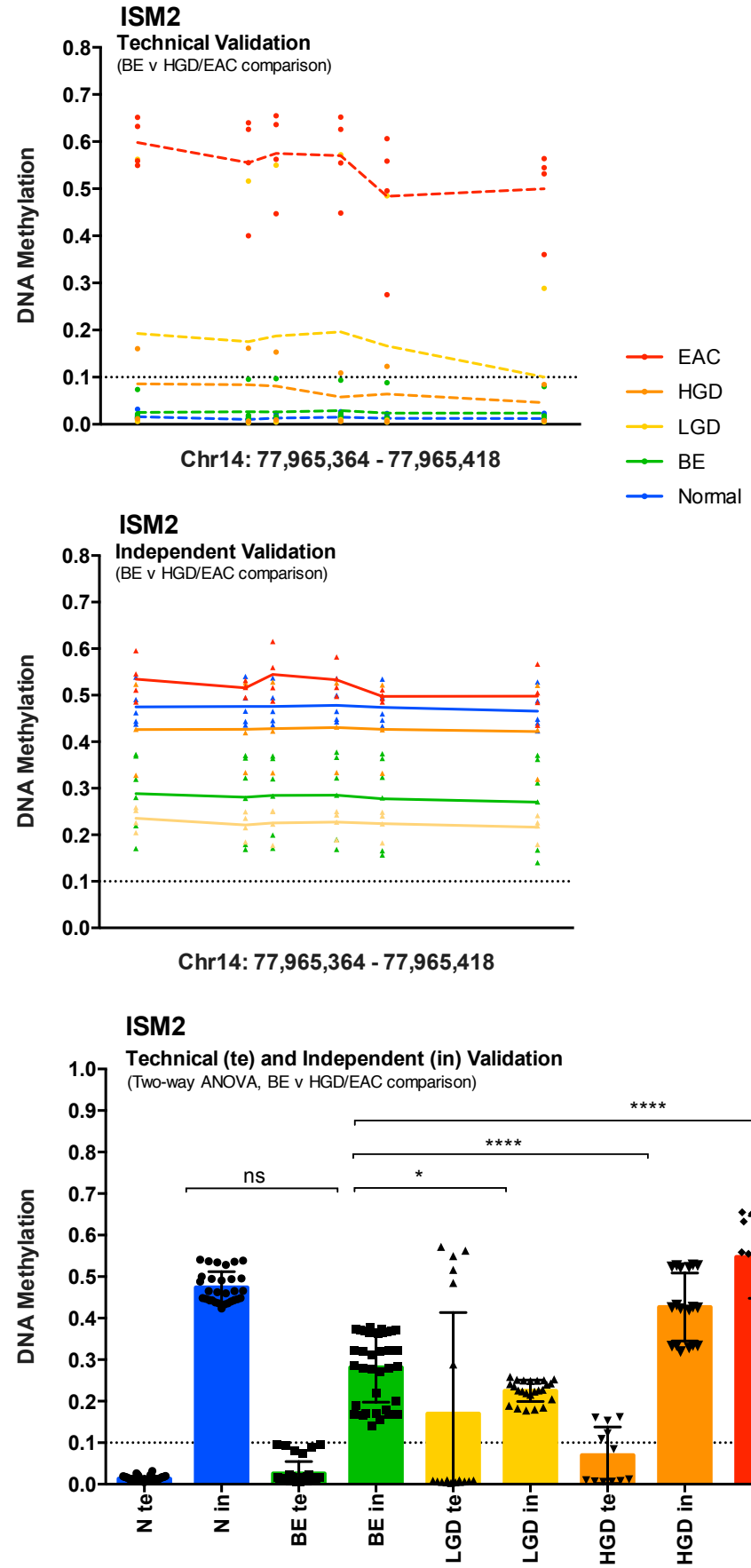
Of the 13 target regions (refer to Table 5-11) for identification of intervention requiring disease (BE v HGD-EAC), there was one incidence of failure of independent cohort to validate technical cohort results (ISM2), hypothesized to be attributable to technical issues with target amplification for the independent cohort; supported by external validation cohort results for this region (Figure 5-8). Technical validation sequencing supported genome-wide methylation profiling with average β (all CpG sites in region): 0.013 ± 0.006 and 0.026 ± 0.029 for N and non-dysplastic BE samples respectively, and 0.388 ± 0.245 for HGD-EAC ($\Delta\beta>0.35$). This was further supported by HM450 data for the 75 N and 19 BE samples in the external validation cohort (average methylation of probes in the selected target region $\beta < 0.05$ in 75 of 75 N and 18 of 19 BE samples), but disparate with independent validation

sequencing results (average β (all CpG sites in region): 0.47 ± 0.04 and 0.28 ± 0.08 for N and non-dysplastic BE samples respectively).

Isthmin (ISM) encodes a protein expressed in skin and mucosal tissues, with proapoptotic, antiangiogenic and antitumorigenic properties. It has been shown to induce apoptosis in endothelial cells, but under different conditions may promote adhesion and survival in these same cells[156, 157]. In 2008, Xiang et al[158] examined ISM and its role in tumor growth, finding that its expression inhibited endothelial cell capillary network formation and significantly suppressed melanoma tumor growth in mice. They found that the suppressed tumor growth was due to inhibition of angiogenesis; ISM overexpression did not affect tumor cell proliferation. At the time, this was ground-breaking as ISM had no known functions. Angiogenesis (the formation of new blood vessels from pre-existing ones) is an important factor in tumor progression and metastases with limitation of angiogenesis a treatment approach emerging across a number of cancer types[159-161]. ISM1 has been shown to inhibit glioma growth by inhibition of angiogenesis; with Yuan et al identifying isthmin as a novel angiogenesis inhibitor for glioma therapy in 2012[162].

My initial findings and supporting validation of the presence of aberrant promoter ISM2 hypermethylation in EAC leads to the hypothesis that this aberrant methylation may lead to downstream silencing of isthmin, allowing angiogenesis to go unchecked, promoting tumor progression and metastasis by the formation of new blood vessels. Failure of independent validation cohort sequencing to support technical cohort sequencing and genome-wide sequencing of the large external validation cohort supports conclusions of technical issues with target amplification of this region for independent sequencing validation.

(A)


ISM2
Technical Validation
(BE v HGD/EAC comparison)
DNA Methylation
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Chr14: 77,965,364 - 77,965,418
EAC
HGD
LGD
BE
Normal
ISM2
Independent Validation
(BE v HGD/EAC comparison)
DNA Methylation
Chr14: 77,965,364 - 77,965,418
ISM2
Technical (te) and Independent (in) Validation
(Two-way ANOVA, BE v HGD/EAC comparison)
DNA Methylation
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
ns
*
****
****
N te
N in
BE te
BE in
LGD te
LGD in
HGD te
HGD in
EAC te
EAC in

(B)

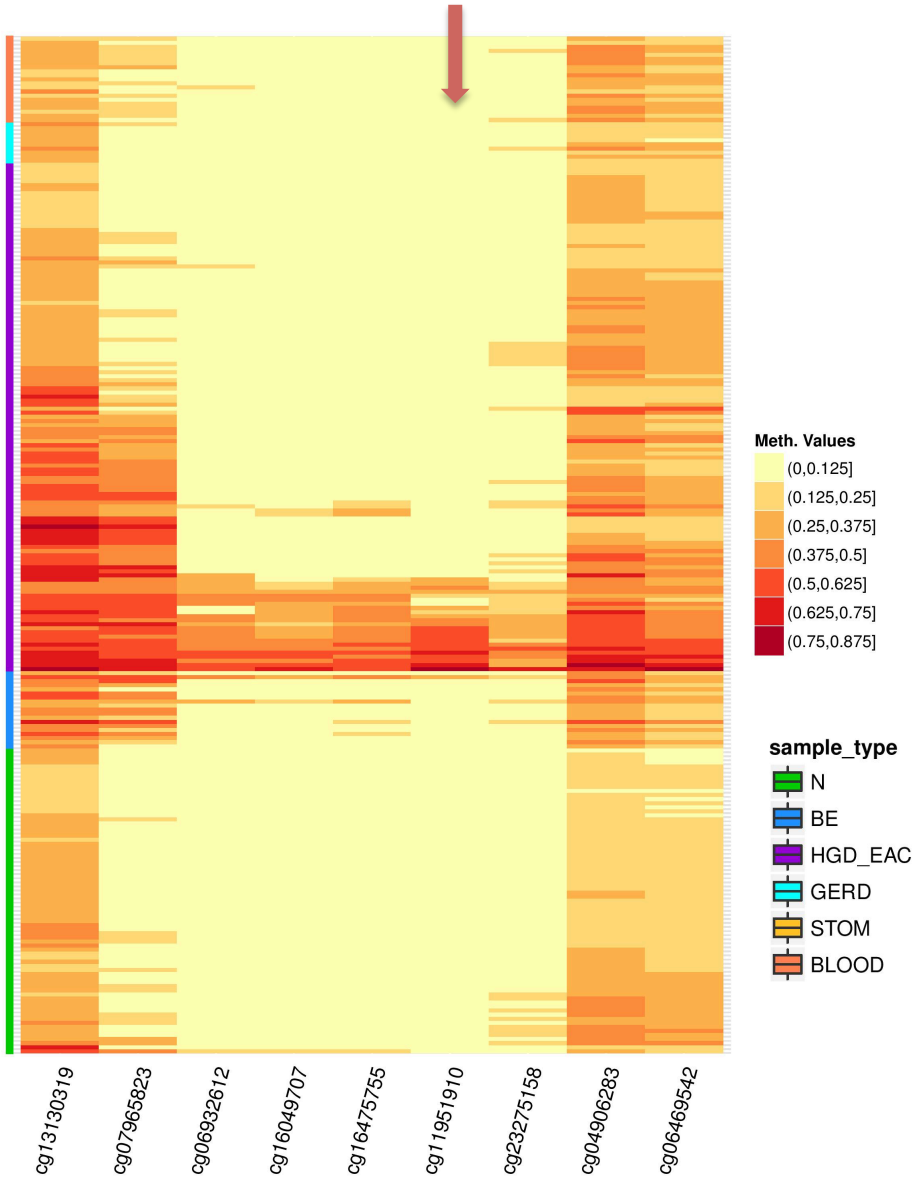


Figure 5-8: Validation of disease-associated differential methylation occurring in the promoter region of Isthmin 2 (ISM2) failed for independent validation cohort sequencing. (A) Unexpected high-level DNA methylation was observed across all disease classes in targeted sequencing of the independent validation cohort. This did not support the EAC associated hypermethylation observed in the technical cohort (B) Genome-wide methylation profiling data for the external validation cohort (250 samples) supports differential methylation observed in the training (HM450) and technical validation (sequencing) cohorts; with hypermethylation only detected in a subset of EAC samples (normal squamous epithelium, Barrett's mucosa, and inflamed epithelium from GERD patients all entirely

unmethylated). Elongated validation region includes 9 probes, Chr14: 77,964,542 – 77,965,870. Targeted amplicon sequencing region (containing six CpG sites, Chr14: 77,965,343 – 77,965,434; indicated by red arrow,) contains probe cg11951910. Heatmap methylation values separated into eight discrete regions 0.000 – 1.000 in increments of 0.125.

Further to the failed sequencing validation of ISM2, there were two incidences (TRANK1, TNFAIP8L3) of methylation above expected baseline in normal and non-dysplastic BE in both technical and independent sequencing validation when compared back to genome-wide methylation profiling (training cohort and external validation cohort, data not shown). Methylation controls (50:50 methylated:unmethylated standard DNA) were run as part of the independent validation cohort to determine any target region amplification bias. TRANK1 and TNFAIP8L3 showed methylation amplification bias of 26.3 ± 0.8% and 15.2 ± 0.7% respectively, showing that amplification prior to targeted sequencing was responsible for skewed methylation results for both these targets. External HM450 data for the 75 normal squamous esophageal and 19 non-dysplastic BE comprising this cohort supported the unmethylated status of this category of samples; supporting hypotheses of amplification bias prior to sequencing. Despite amplification bias, targeted sequencing results were still able to return statistically significant $\Delta\beta$ (BE v EAC, $p<0.0001$) for both TRANK1 and TNFAIP8L3 target regions.

Despite amplification bias of TRANK1 and TNFAIP8L3, both regions are interesting targets. In 2015, in a genome-wide approach to link clinical outcome with genotype in breast cancer, Pongor et al identified Tetratricopeptide Repeat and Ankyrin Repeat Containing 1 (TRANK1) as one of the top 10 driver oncogenes, citing 5.3% mutational prevalence[163]. In the case of esophageal adenocarcinoma, an inflammation-induced stem cell origin rather than a driver mutation origin is favored; but this is not to say that accumulation of mutations during tumorigenesis is not occurring; indeed, these mutations may be required for tumor progression and metastasis to occur.

The Tumor Necrosis Factor, Alpha-Induced Protein 8 (TNFAIP8) family, only identified in 2008[164, 165] encodes lipid transfer proteins, capturing and shuttling phosphatidylinositol 4,5-bisphosphate and phosphatidylinositol 3,4,5-triphosphate across the plasma membrane[166]. More than half of all human cancers have hyperactivation of phosphoinositide signaling and marked up-regulation of the TNFAIP8L2 (TIPE2), yet lipid signaling in cancer cells is not yet fully understood[167-169]. In the TNFAIP8 family, most is known about TNFAIP8L2 (TIPE2), which regulates both inflammation and carcinogenesis[170], whereas the role of TNFAIP8L3 (also known as TIPE3) has remained largely unknown until very recently[167]. Fayngerts et al (2014) showed that TNFAIP8L3 knockout diminished tumorigenesis and enforced expression enhances oncogenesis in lung, bladder and colorectal carcinoma cell lines[167]. The oncogenic role of TNFAIP8L3 has lead to its proposal as a therapeutic target for malignant disease treatment; support for this has yet to be published.

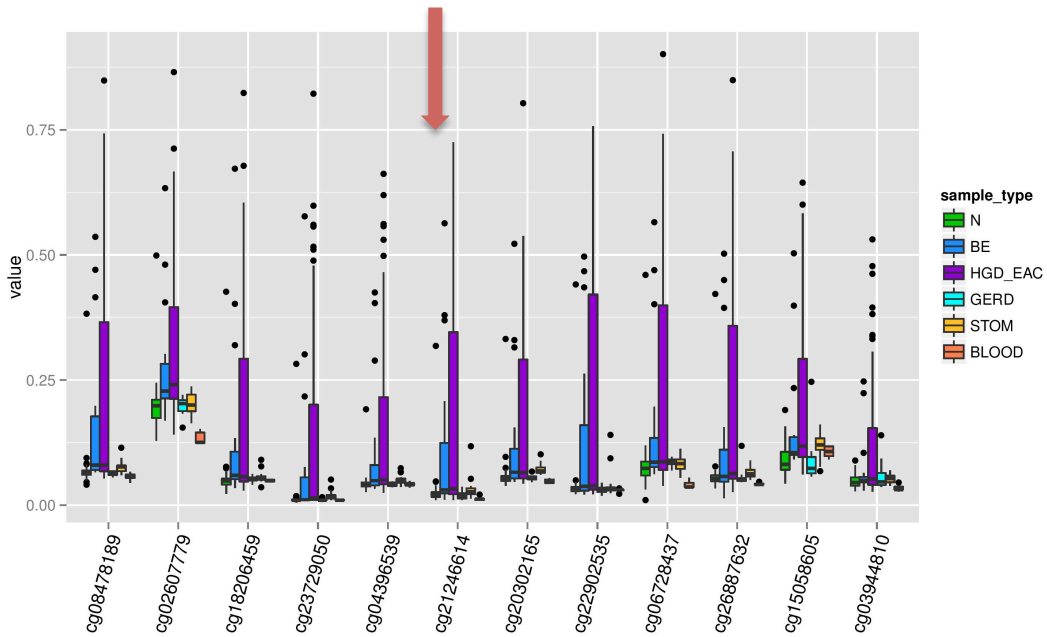
5.3.5 External differential methylation validation

Differential methylation of selected target regions, elongated to include flanking regions on either side, were validated using genome-wide methylation profiling data from an external validation cohort, consisting of 250 samples from 154 patients (125 EAC, 19 BE (8 from BE patients, 11 adjacent BE from EAC patients), 85 non-tumor squamous esophagus and 21 normal stomach samples (Table 3-2). Average methylation beta-values at each probe were determined to produce barplots and heatmaps for all target regions. Examples of validation for KLF7, LRRC43 and VANGL2 are given in Figure 5-9, Figure 5-10 and Figure 5-11 respectively. External validation results support internal genome-wide methylation profiling, as well as internal validation by targeted amplicon sequencing.

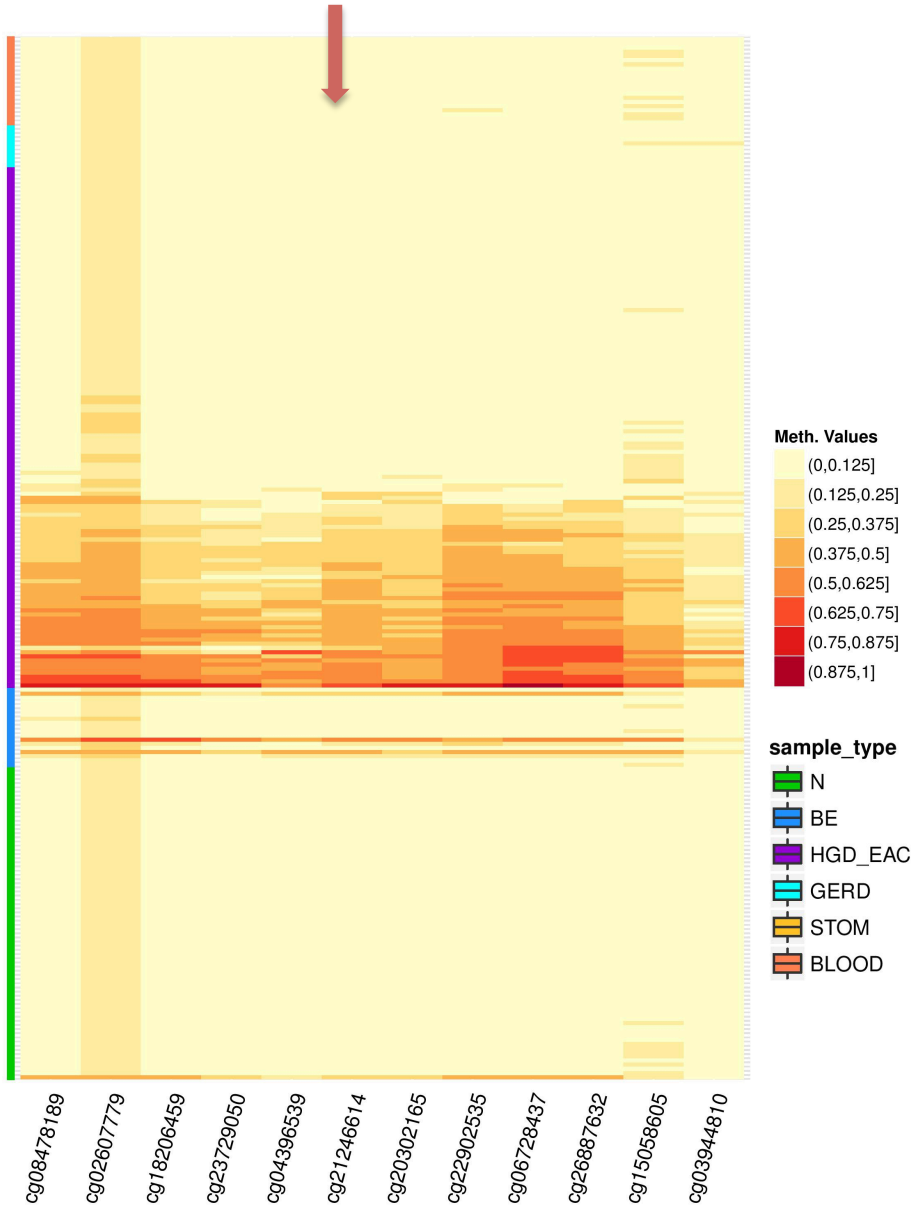
Kruppel-like factor 7 (KLF7) is a member of the Kruppel-like family of DNA-binding transcription factors, regulating cell proliferation, differentiation and survival. Members in this family contain three C_2H_2 zinc fingers at the extreme carboxyl end, aiding in binding to GC rich sites. Expression of KLF4

and KLF7 have been found to be critical for decisions between proliferation and cell cycle arrest and differentiation[171]. The proteins encoded by this family of genes have been implicated in type 2 diabetes progression[172], as well as recently identified as a tumor suppressor genes involved in carcinogenesis and metastasis in a variety of cancer types. A flurry of publications in 2015 associated various KLF family members with carcinogenic progression: KLF4 in myeloid leukemia, breast and esophageal squamous cell carcinoma[173-175], KLF2 in pancreatic ductal adenocarcinoma[176], KLF8 in colorectal cancer[177] and KLF17 in breast and esophageal cancer[178, 179]. The role of epigenetic silencing of this family of DNA-binding transcription factors and regulation of metastasis has been examined by Sacheva et al (2015), who published evidence that the epigenetic silencing of KLF3 by promoter hypermethylation in primary mouse and human metastatic sarcoma increases pro-metastatic miR-182[180]. My findings of aberrant promoter hypermethylation in KLF7 in dysplastic disease and esophageal adenocarcinoma are therefore not surprising; however the involvement of this particular family member in EAC carcinogenesis is novel and further studies to elucidate mechanisms of involvement are warranted.

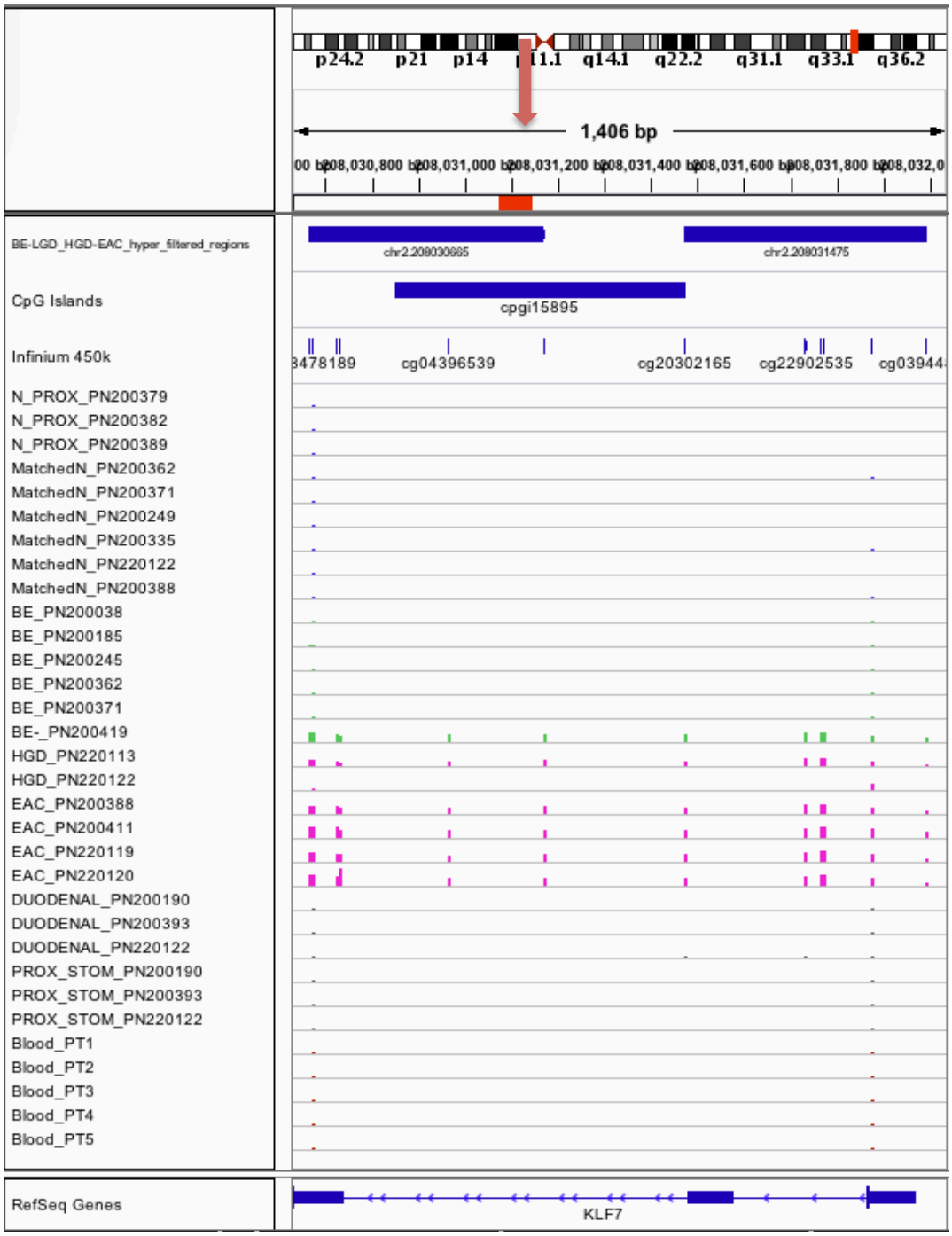
(A)



(B)


Meth. Values
(0,0.125]
(0.125,0.25]
(0.25,0.375]
(0.375,0.5]
(0.5,0.625]
(0.625,0.75]
(0.75,0.875]
(0.875,1]
sample_type
N
BE
HGD_EAC
GERD
STOM
BLOOD
cg08478189
cg02607779
cg18206459
cg23729050
cg04396539
cg21246614
cg20302165
cg22902535
cg06728437
cg26887632
cg15058605
cg03944810

(C)


p24.2
p21
p14
q14.1
q22.2
q31.1
q33.1
q36.2
1,406 bp
BE-LGD_HGD-EAC_hyper_filtered_regions
chr2:208030665
chr2:208031475
CpG Islands
cpgi15895
Infinium 450k
cg04396539
cg20302165
cg22902535
N_PROX_PN200379
N_PROX_PN200382
N_PROX_PN200389
MatchedN_PN200362
MatchedN_PN200371
MatchedN_PN200249
MatchedN_PN200335
MatchedN_PN220122
MatchedN_PN200388
BE_PN200038
BE_PN200185
BE_PN200245
BE_PN200362
BE_PN200371
BE-_PN200419
HGD_PN220113
HGD_PN220122
EAC_PN200388
EAC_PN200411
EAC_PN220119
EAC_PN220120
DUODENAL_PN200190
DUODENAL_PN200393
DUODENAL_PN220122
PROX_STOM_PN200190
PROX_STOM_PN200393
PROX_STOM_PN220122
Blood_PT1
Blood_PT2
Blood_PT3
Blood_PT4
Blood_PT5
RefSeq Genes
KLF7

(D)

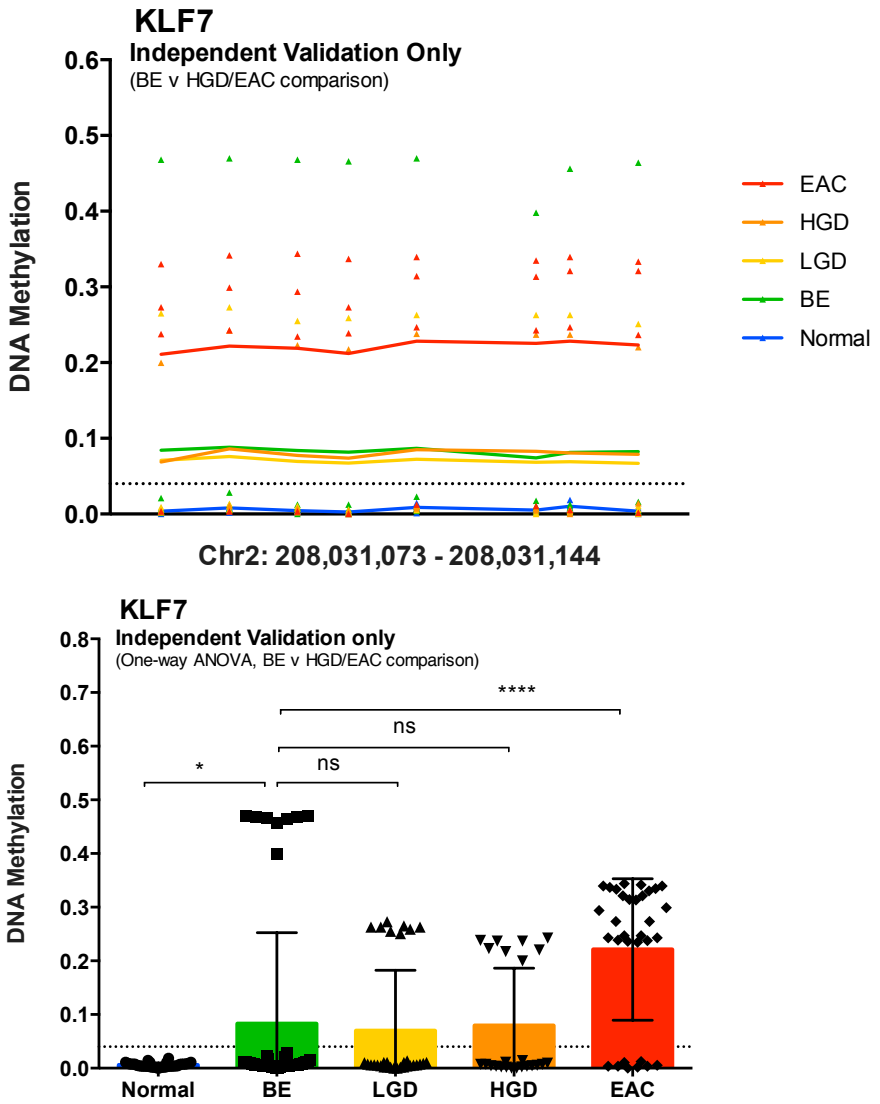


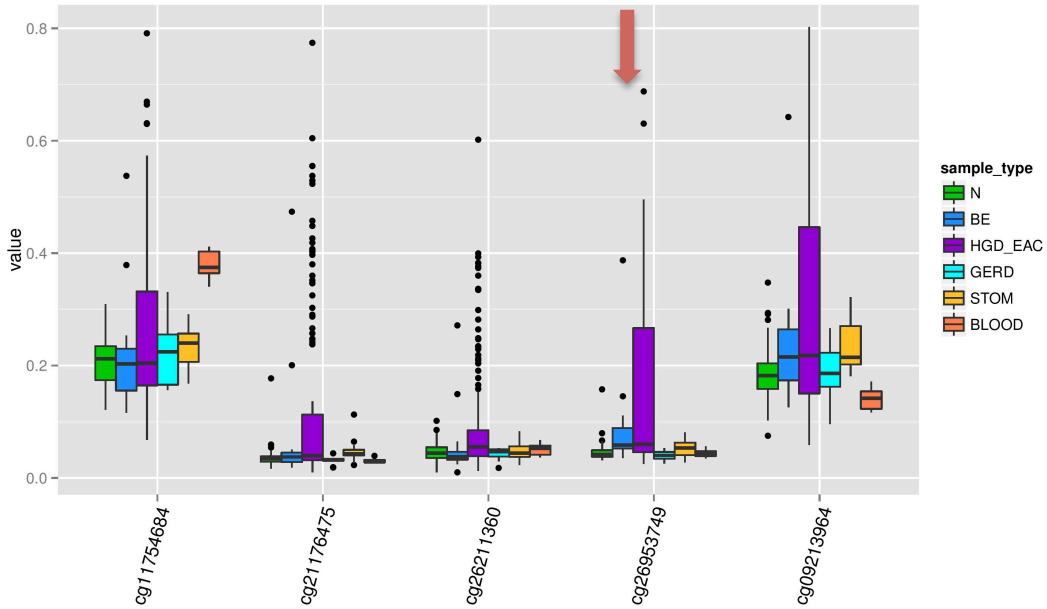
Figure 5-9: Validation of disease-associated differential methylation occurring in the promoter region of Kruppel-like factor 7 (ubiquitous) (KLF7), identified using the clinically relevant BE v HGD-EAC comparison for identification of intervention-requiring disease. Elongated validation region examined includes 12 probes, Chr2: 208,030,666 – 208,031,996. Targeted amplicon sequencing region (eight CpG sites, Chr2: 208,031,073-208,031,144), indicated by red arrow, lies between probes cg04396539 and cg21246614. Barplot (A) and heatmap (B) show External Validation Cohort results. Genome browser (C) shows HM450 data for a subset of the Training Cohort with >90% tissue sample homogeneity. DNA methylation evaluation by targeted amplicon sequencing (D) shows results for Independent Validation Cohort. External validation (A) and (B) supports internal genome-wide

methylation profiling of the Training Cohort (C), as well as internal validation by targeted amplicon sequencing of the Independent Validation Cohort (D). There is evidence of aberrant methylation in a subset of dysplastic disease and EAC samples, with external validation supporting our observed bimodal methylation of advanced disease samples. Intriguing for this target region was increased methylation in occasional BE samples. There was no supporting evidence of dysplastic progression of these patients (Training or Independent Validation Cohorts) during the course of this study. Follow-up clinical data was not available for the External Validation Cohort. (B) Heatmap methylation values separated into eight discrete regions 0.000 – 1.000 in increments of 0.125. Heatmap sample classification color scheme as for barplot (A). (C) Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

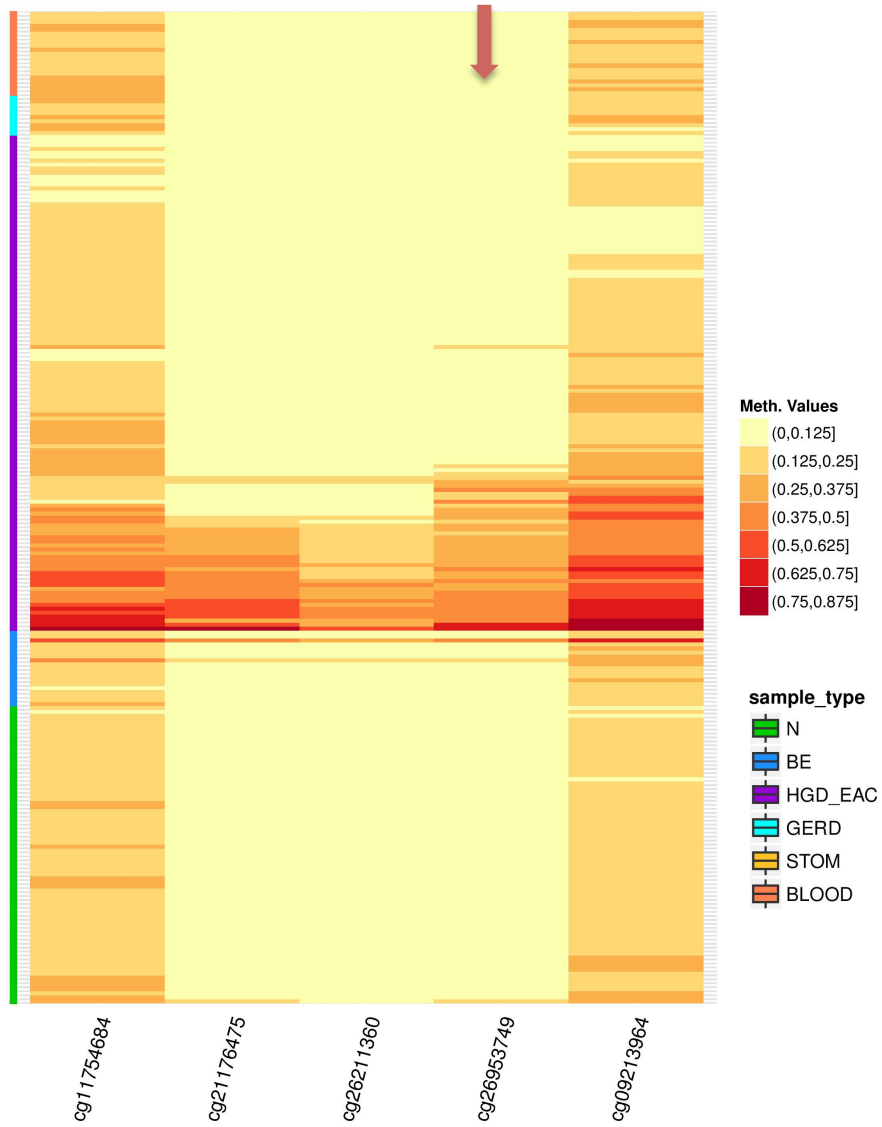
Leucine rich repeat containing 43 (LRRC43) is a member of the leucine-rich repeat (LRR) superfamily, involved in a diverse range of cell-cell and cell-extracellular matrix interactions such as adhesion, target recognition and receptor-ligand binding[181]. The human protein atlas (www.proteinatlas.org) shows evidence of LRRC43 expression (high level antibody staining) across a number of cancers including colorectal, glioma, melanoma and head and neck cancer[182, 183]. Overexpression of LRR superfamily member, leucine rich repeat containing G-protein coupled receptor 5 (LGR5), an intestinal stem cell marker, has been observed in hepatocellular carcinoma with β-catenin mutation (LGR5 is a downstream target gene of the Wnt signaling pathway)[184]. Furthermore, LGR5 has also been found to have prognostic significance in gastric carcinoma (expression associated with poor survival)[185], but only in gastric cancers with nuclear β-catenin positivity. Interestingly, increased LGR expression did NOT have any effect on growth or migration of gastric carcinoma cells[186]. LRRC15 has been found to be frequently overexpressed in breast carcinoma [187-189] as well as androgen-independent prostate cancer[190] and is associated with aggressive, high-grade tumors in both cases. As a result of this association, O'Prey et al sought to investigate the role of LRRC15 on adenoviral delivery of tumor

suppressor 53; which was found to repress cell death due to adenoviral p53 by impeding adenoviral infection[191]. Thus, despite some research into the LRR superfamily and its role in carcinogenesis in a number of cancer types; functional roles of specific family members are as yet unclear. We can hypothesize that LRR family members (such as LRRC43, identified here to exhibit EAC-associated aberrant promoter hypermethylation in genome-wide methylation profiling studies, validated by targeted sequencing (Figure 5-10)), may be responsible for mediating nuclear import of oncoprotein (such as LRCC59-PP2A interaction in prostate cancer[192]); however further functional studies would be necessary to confirm this hypothesis.

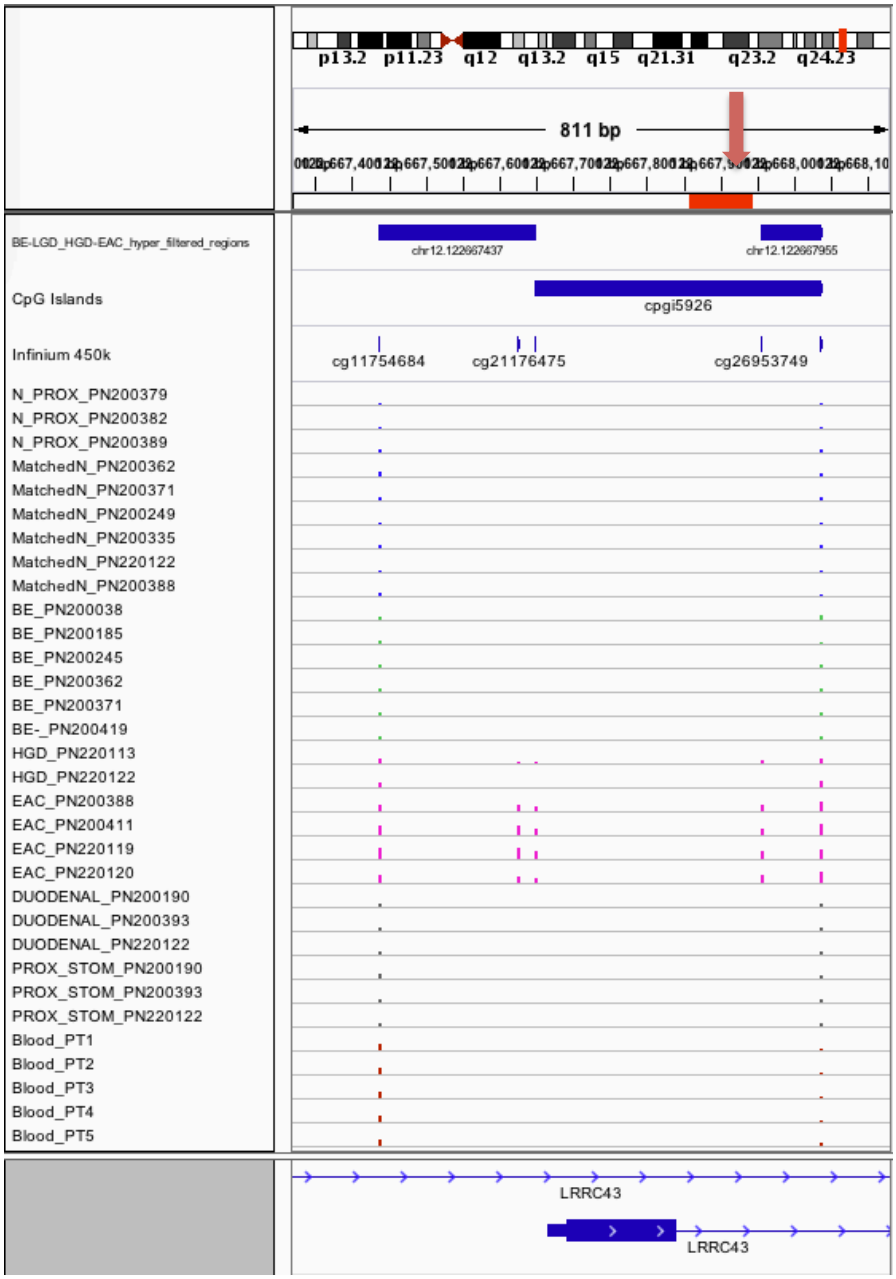
(A)



(B)


Meth. Values
(0,0.125]
(0.125,0.25]
(0.25,0.375]
(0.375,0.5]
(0.5,0.625]
(0.625,0.75]
(0.75,0.875]
sample_type
N
BE
HGD_EAC
GERD
STOM
BLOOD
cg11754684
cg21176475
cg26211360
cg26953749
cg09213964

(C)


p13.2 p11.23 q12 q13.2 q15 q21.31 q23.2 q24.23
811 bp
BE-LGD_HGD-EAC_hyper_filtered_regions
chr12:122667437
chr12:122667955
CpG Islands
cpgi5926
Infinium 450k
cg11754684
cg21176475
cg26953749
N_PROX_PN200379
N_PROX_PN200382
N_PROX_PN200389
MatchedN_PN200362
MatchedN_PN200371
MatchedN_PN200249
MatchedN_PN200335
MatchedN_PN220122
MatchedN_PN200388
BE_PN200038
BE_PN200185
BE_PN200245
BE_PN200362
BE_PN200371
BE-_PN200419
HGD_PN220113
HGD_PN220122
EAC_PN200388
EAC_PN200411
EAC_PN220119
EAC_PN220120
DUODENAL_PN200190
DUODENAL_PN200393
DUODENAL_PN220122
PROX_STOM_PN200190
PROX_STOM_PN200393
PROX_STOM_PN220122
Blood_PT1
Blood_PT2
Blood_PT3
Blood_PT4
Blood_PT5
LRRC43
LRRC43

(D)

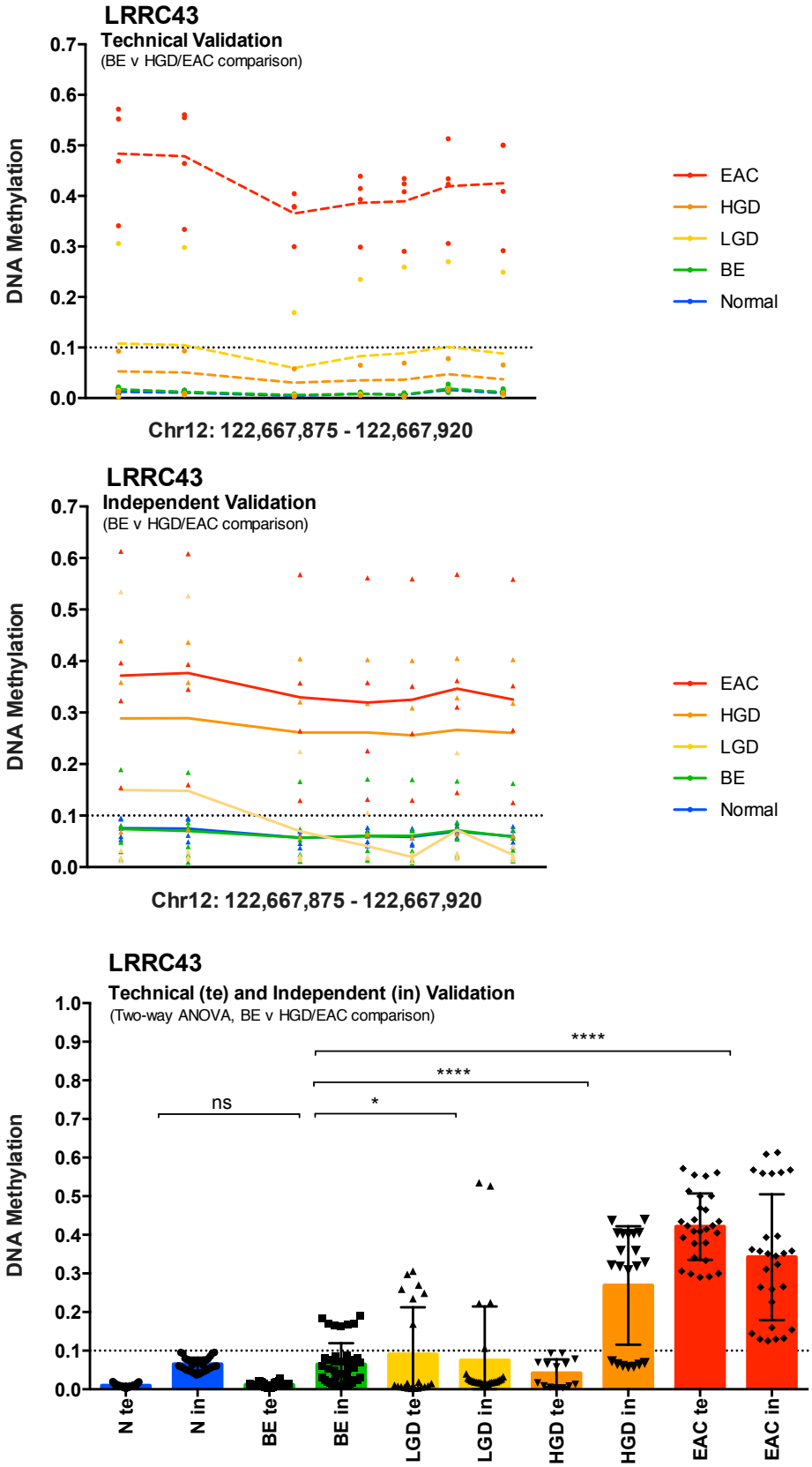


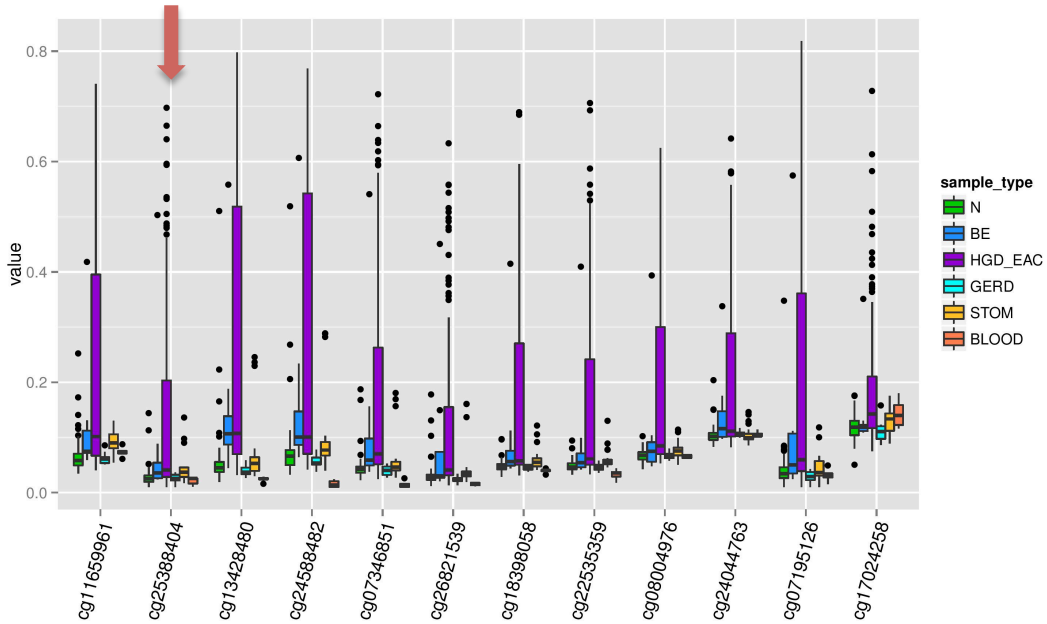
Figure 5-10: Validation of disease-associated differential methylation occurring in the promoter region of leucine rich repeat containing 43

(LRRC43), identified using the clinically relevant BE v HGD-EAC comparison for identification of intervention-requiring disease. Elongated validation region examined includes 5 probes, Chr12: 122,667,438 – 122,668,038. Targeted amplicon sequencing region (seven CpG sites, Chr12: 122,667,942 – 122,667,857), indicated by red arrow, lies next to probe cg26953749. Barplot (A) and heatmap (B) show External Validation Cohort results. Genome browser (C) shows HM450 data for a subset of the Training Cohort with >90% tissue sample homogeneity. DNA methylation evaluation by targeted amplicon sequencing (D) shows results for both the Technical and Independent Validation Cohorts. External validation (A) and (B) support internal genome-wide methylation profiling of the Training Cohort (C), as well as internal validation by targeted amplicon sequencing of the Technical and Independent Validation Cohorts (D). There is evidence of aberrant methylation in a subset of dysplastic disease and EAC samples, with external validation supporting our observed bimodal methylation of advanced disease samples. (B) Heatmap methylation values separated into eight discrete regions 0.000 – 1.000 in increments of 0.125. Heatmap sample classification color scheme as for barplot (A). (C) Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

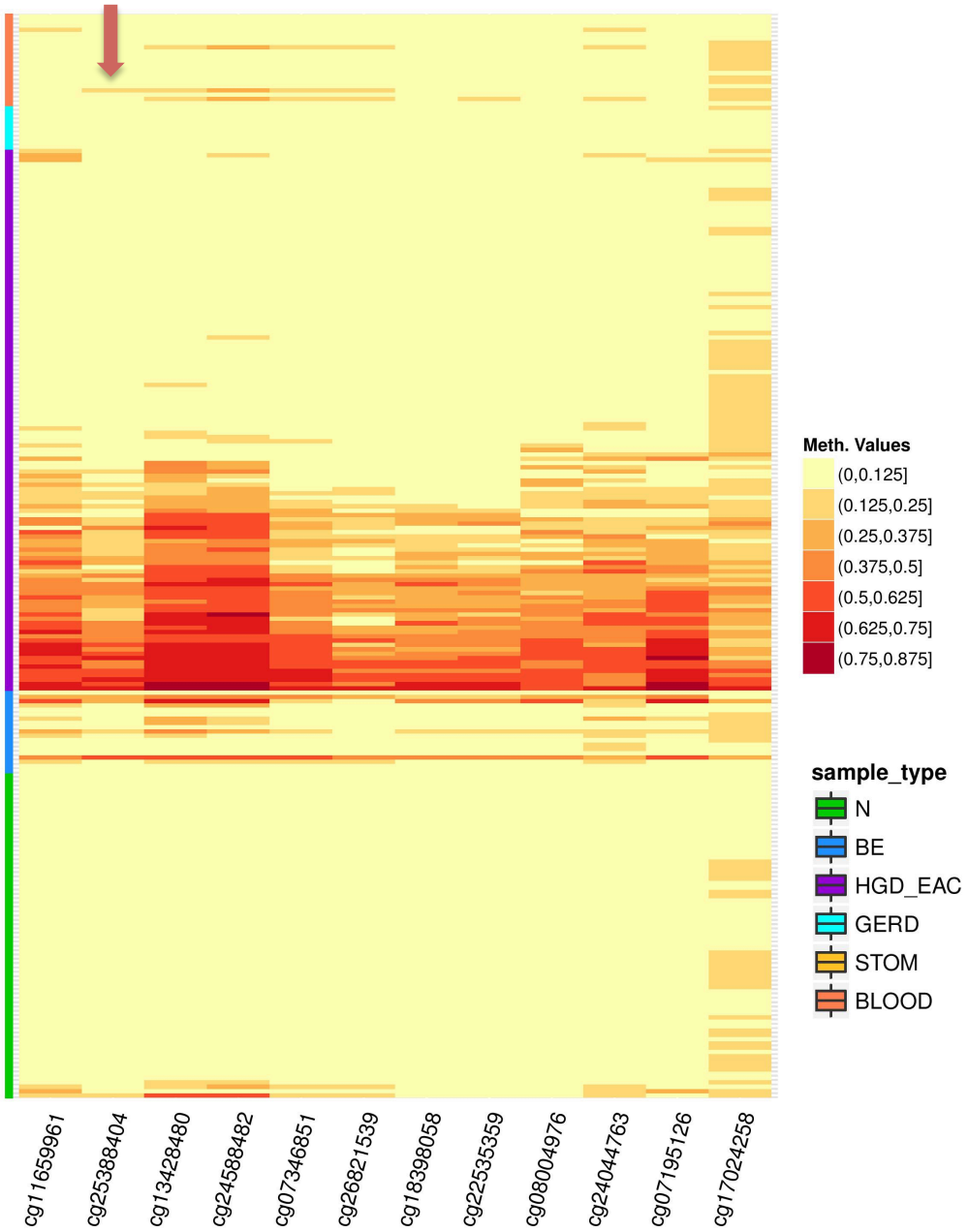
Vang-like 2 (VANGL2) is a developmental membrane protein that plays a role in regulation of cell polarity, most notable for its involvement in neural tube defects occurring during the first trimester of pregnancy[193-195]. However, a developing role for VANGL2 in cancer has recently begun to be elucidated. The reactivation of embryonic development pathways to promote aggressive cell behaviour is not uncommon in cancer. Initial data mining efforts by Hatakeyama et al (2014) suggest that VANGL1 and VANGL2 are dysregulated in human cancers, and include the observation that estrogen receptor (ER)-positive breast cancer patients with elevated VANGL1 expression correlate with worse overall survival[196]. Aberrant promoter methylation in a subset of dysplastic disease and EAC suggests that VANGL2 may also be involved in esophageal carcinogenesis, a novel finding in the EAC field. Reactivation of this embryonic developmental pathway

increases invasion and metastasis[197], therefore we could hypothesize that the subset of patients with aberrantly methylated VANGL2 may represent more aggressive cancers with worse survival. Comprehensive analysis with associated long-term survival data would be required to test this hypothesis.

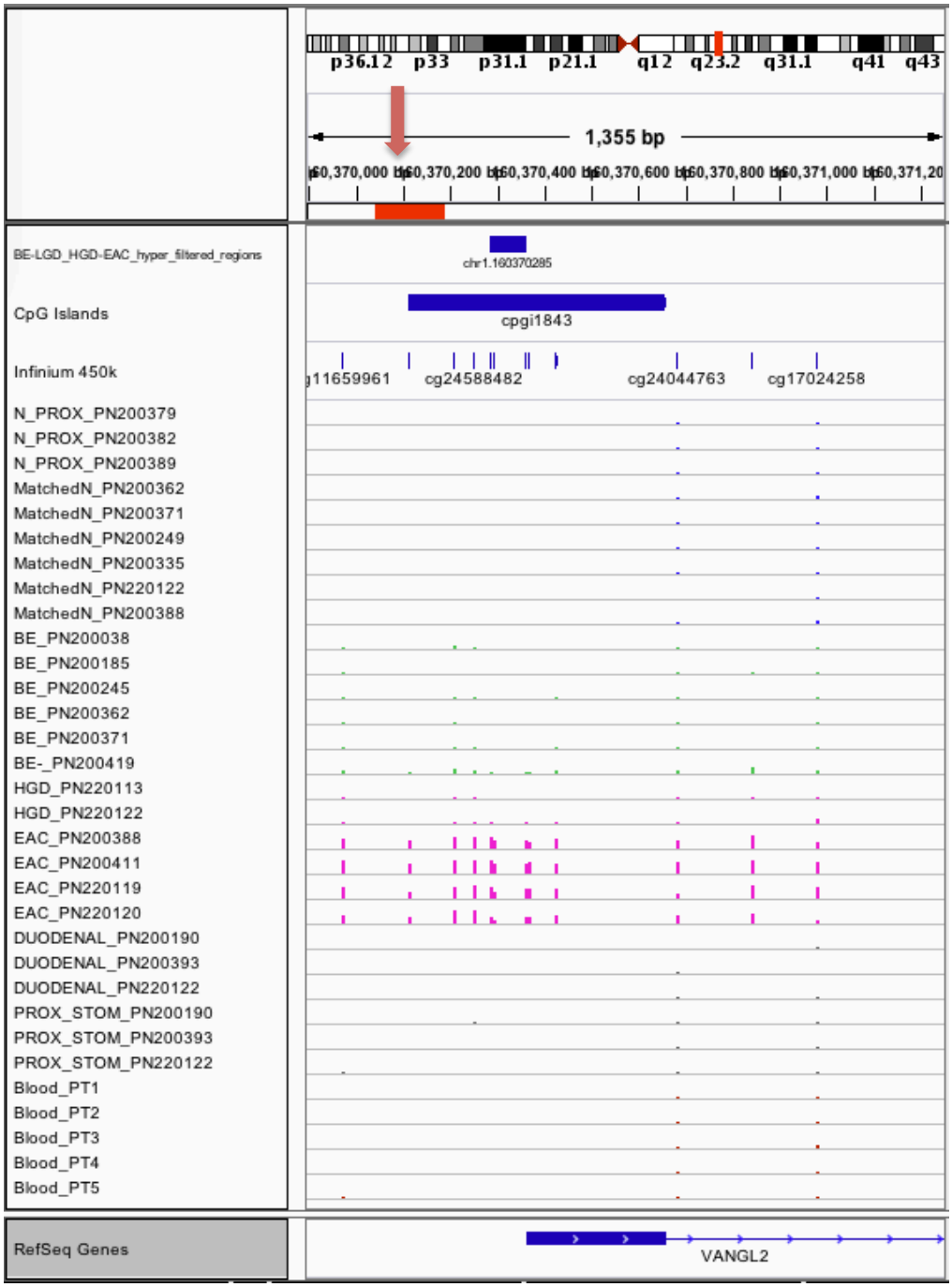
(A)



(B)


Meth. Values
(0,0.125]
(0.125,0.25]
(0.25,0.375]
(0.375,0.5]
(0.5,0.625]
(0.625,0.75]
(0.75,0.875]
sample_type
N
BE
HGD_EAC
GERD
STOM
BLOOD
cg11659961
cg25388404
cg13428480
cg24588482
cg07346851
cg26821539
cg18398058
cg22535359
cg08004976
cg24044763
cg07195126
cg17024258

(C)


p36.12 p33 p31.1 p21.1 q12 q23.2 q31.1 q41 q43
1,355 bp
BE-LGD_HGD-EAC_hyper_filtered_regions
chr1:160370285
CpG Islands
cpgi1843
Infinium 450k
cg11659961
cg24588482
cg24044763
cg17024258
N_PROX_PN200379
N_PROX_PN200382
N_PROX_PN200389
MatchedN_PN200362
MatchedN_PN200371
MatchedN_PN200249
MatchedN_PN200335
MatchedN_PN220122
MatchedN_PN200388
BE_PN200038
BE_PN200185
BE_PN200245
BE_PN200362
BE_PN200371
BE-_PN200419
HGD_PN220113
HGD_PN220122
EAC_PN200388
EAC_PN200411
EAC_PN220119
EAC_PN220120
DUODENAL_PN200190
DUODENAL_PN200393
DUODENAL_PN220122
PROX_STOM_PN200190
PROX_STOM_PN200393
PROX_STOM_PN220122
Blood_PT1
Blood_PT2
Blood_PT3
Blood_PT4
Blood_PT5
RefSeq Genes
VANGL2

(D)

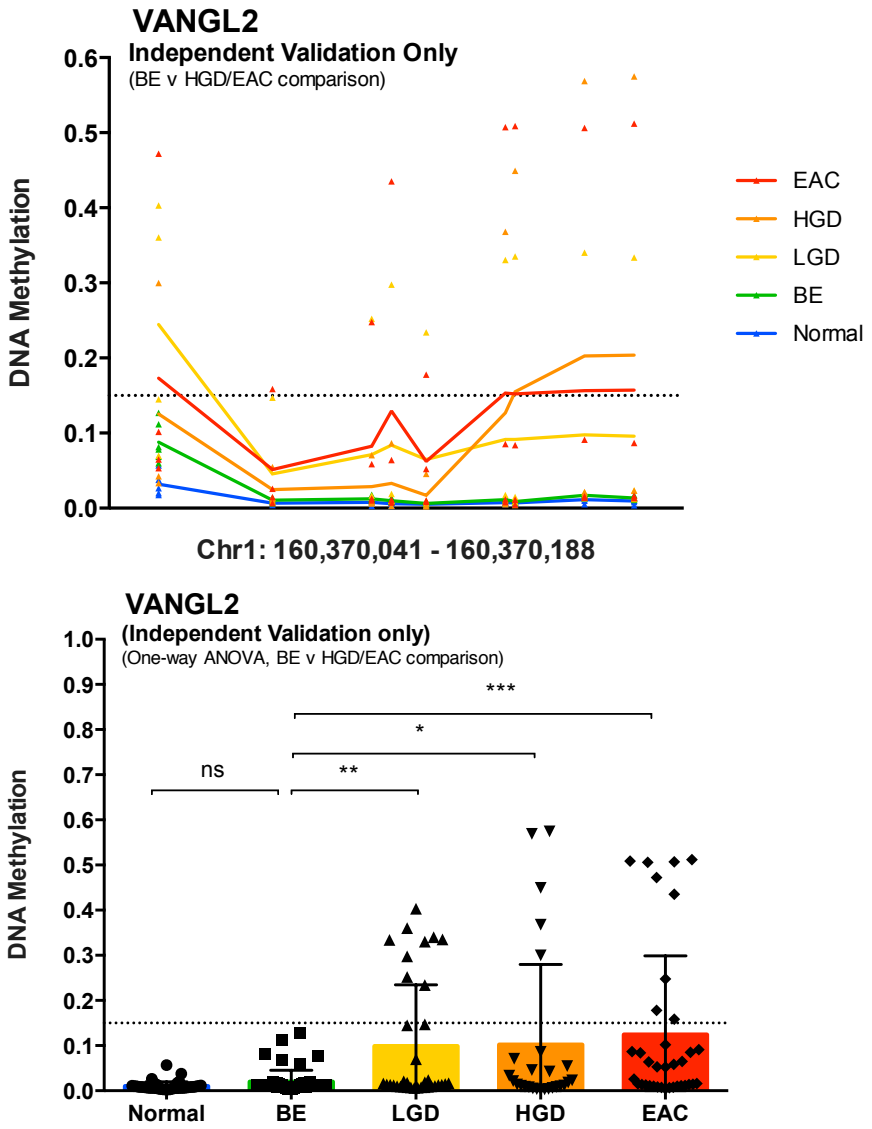


Figure 5-11: Validation of disease-associated differential methylation occurring in the promoter region of vang-like 2 (VANGL2), identified using the clinically relevant BE v HGD-EAC comparison for identification of intervention-requiring disease. Elongated validation region examined includes 12 probes, Chr1: 160,369,972 – 160,370,983. Targeted amplicon sequencing region (nine CpG sites, Chr1: 160,370,041 – 160,370,188), indicated by red arrow, includes probe cg25388404. Barplot (A) and heatmap (B) show External Validation Cohort results. Genome browser (C) shows HM450 data for a subset of the Training Cohort with >90% tissue sample homogeneity. DNA methylation evaluation by targeted amplicon sequencing (D) shows results for the Independent Validation Cohort. External validation

(A) and (B) support internal genome-wide methylation profiling of the Training Cohort (C), as well as internal validation by targeted amplicon sequencing of the Independent Validation Cohort (D). There is evidence of aberrant methylation in a subset of dysplastic disease and EAC samples, with external validation supporting our observed bimodal methylation of advanced disease samples. (B) Heatmap methylation values separated into eight discrete regions 0.000 – 1.000 in increments of 0.125. Heatmap sample classification color scheme as for barplot (A). (C) Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

5.3.5.1 Promoter hypermethylation of DNA repair gene MGMT

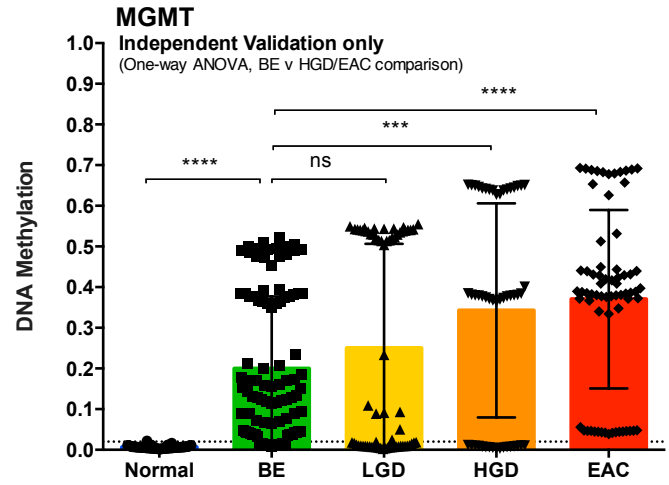
O-6-methylguanine-DNA methyltransferase (MGMT) is a DNA repair gene, crucial for genome stability. MGMT repairs naturally occurring carcinogenic and mutagenic O6-methylguanine back to guanine, preventing mismatch and errors during replication and transcription[198]. The majority of cancers with DNA repair deficiency are epigenetically rather than mutationally regulated, for example, Halford et al reported just 6 of 113 sequential colorectal cancer samples evaluated had missense MGMT mutation; most having reduced MGMT expression due to promoter hypermethyation[199]. Aberrant MGMT promoter hypermethylation plays a significant role in the multistep process of Barrett's carcinogenesis, with 71-79% of EAC reported to be deficient in MGMT[93, 200].

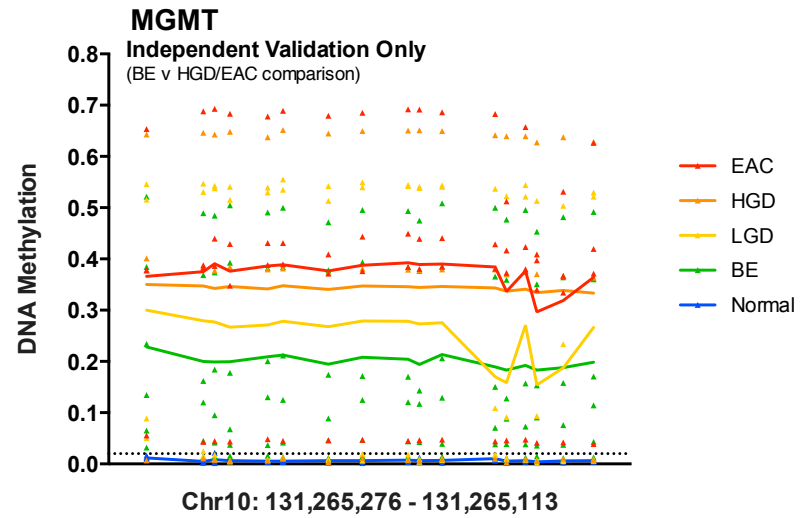
As explained in a recent review by Bernstein et al (2105)[153], epigenetic silencing of MGMT would likely increase mutation rates; one or more of which may provide the cell with selective advantage. Carried on, this continued presence of epigenetically repressed MGMT would generate further mutations, some of which could produce a tumor. From this, we could hypothesize that the subset of non-dysplastic Barrett's mucosa with increased promoter MGMT methylation would be more susceptible to malignant change, although it would not be possible to predict which of these patients will progress using this information alone. Thus, epigenetic MGMT repression, in co-ordination with repression of other DNA repair genes could

be one avenue for identification of non-dysplastic Barrett's mucosa with malignant-change potential. To this end, Jiang et al[152] evaluated mRNA expression of 27 repair genes, in astrocytoma compared to normal brain tissue and found 13, including MGMT, to be significantly down-regulated in in any grade of astrocytoma, including low-grade lesions. There is some evidence of MGMT methylation association with IDH1 mutation, for example 81.3% of IDH1-mutated glioma has methylated MGMT compared non-IDH1 mutated tumors (in which 58.3% were reported to be MGMT methylated)[201]; discussed further in Section 7.3.2.1 IDH1/2 and TET2 mutation in the esophagus.

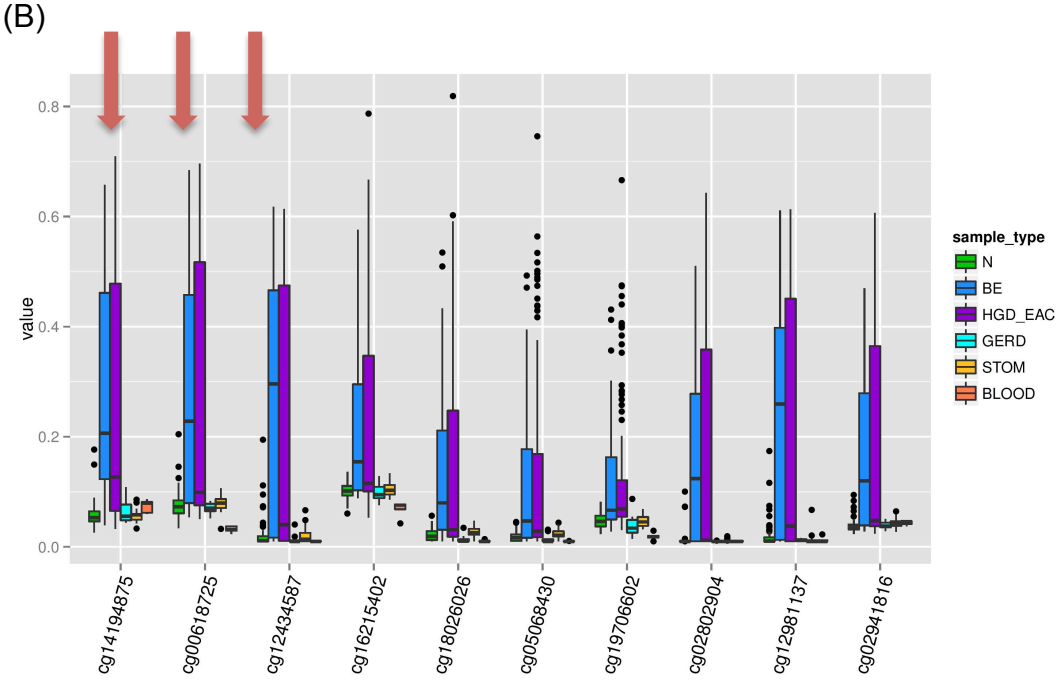
Promoter hypermethylation of MGMT, reported to have diagnostic and prognostic utility in esophageal adenocarcinoma and other cancers[93, 152, 153] was validated here to be present in not only subsets of EAC and dysplastic BE mucosa, but also a subset of non-dysplastic BE mucosa. Methylation bimodality (hypermethylation in only a subset of any disease class, with a subset entirely unmethylated) is observed, in sequencing results (Figure 5-12(A)) as well as HM450 external validation (Figure 5-12(C)). MGMT promoter methylation in non-dysplastic Barrett's mucosa as well as dysplasia and adenocarcinoma is supported by external validation (Figure 5-12) and to some extent, reports in literature. Kuester et al[93] reported MGMT hypermethylation in 21.4% normal esophageal mucosa (n=28), 88.9% of non-dysplastic Barrett's metaplasia (n=27) and 78.9% of EAC (n=47) using methylation specific PCR. For our external validation cohort, MGMT promoter hypermethylation (region as defined for sequencing validation, cut-off $\beta>0.10$) was observed in approximately 3% of normal esophageal mucosa (n=75), 78% of Barrett's mucosa (n=18) and 54% of EAC (n=125).

(A)

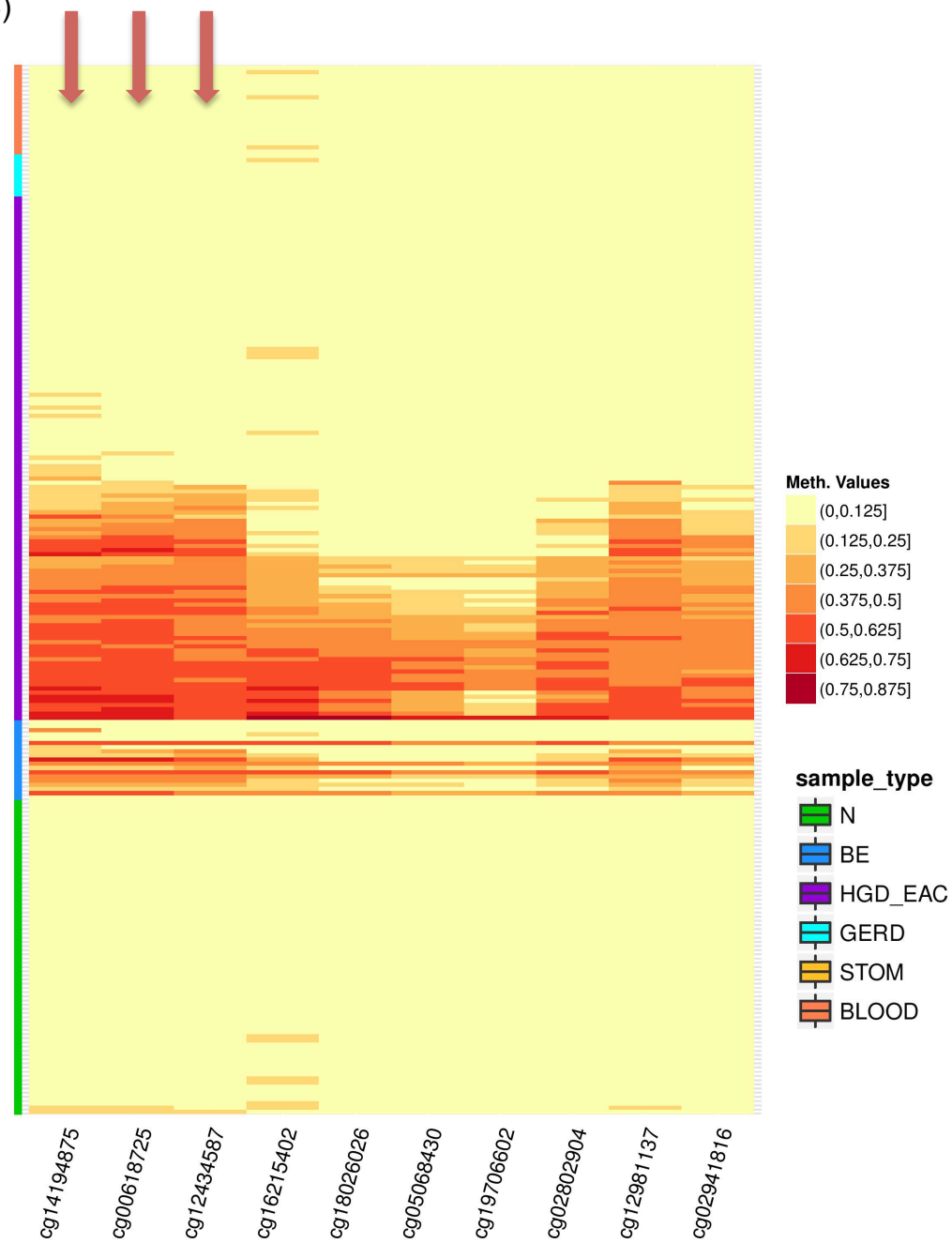

MGMT
Independent Validation only
(One-way ANOVA, BE v HGD/EAC comparison)
DNA Methylation
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
****
ns
***
****
Normal
BE
LGD
HGD
EAC


MGMT
Independent Validation Only
(BE v HGD/EAC comparison)
DNA Methylation
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
EAC
HGD
LGD
BE
Normal
Chr10: 131,265,276 - 131,265,113

(B)


value
0.8
0.6
0.4
0.2
0.0
cg14194875
cg00618725
cg12434587
cg16215402
cg18026026
cg05068430
cg19706602
cg02802904
cg12981137
cg02941816
sample_type
N
BE
HGD_EAC
GERD
STOM
BLOOD

(C)


Meth. Values
(0,0.125]
(0.125,0.25]
(0.25,0.375]
(0.375,0.5]
(0.5,0.625]
(0.625,0.75]
(0.75,0.875]
sample_type
N
BE
HGD_EAC
GERD
STOM
BLOOD
cg14194875
cg00618725
cg12434587
cg16215402
cg18026026
cg05068430
cg19706602
cg02802904
cg12981137
cg02941816

(D)

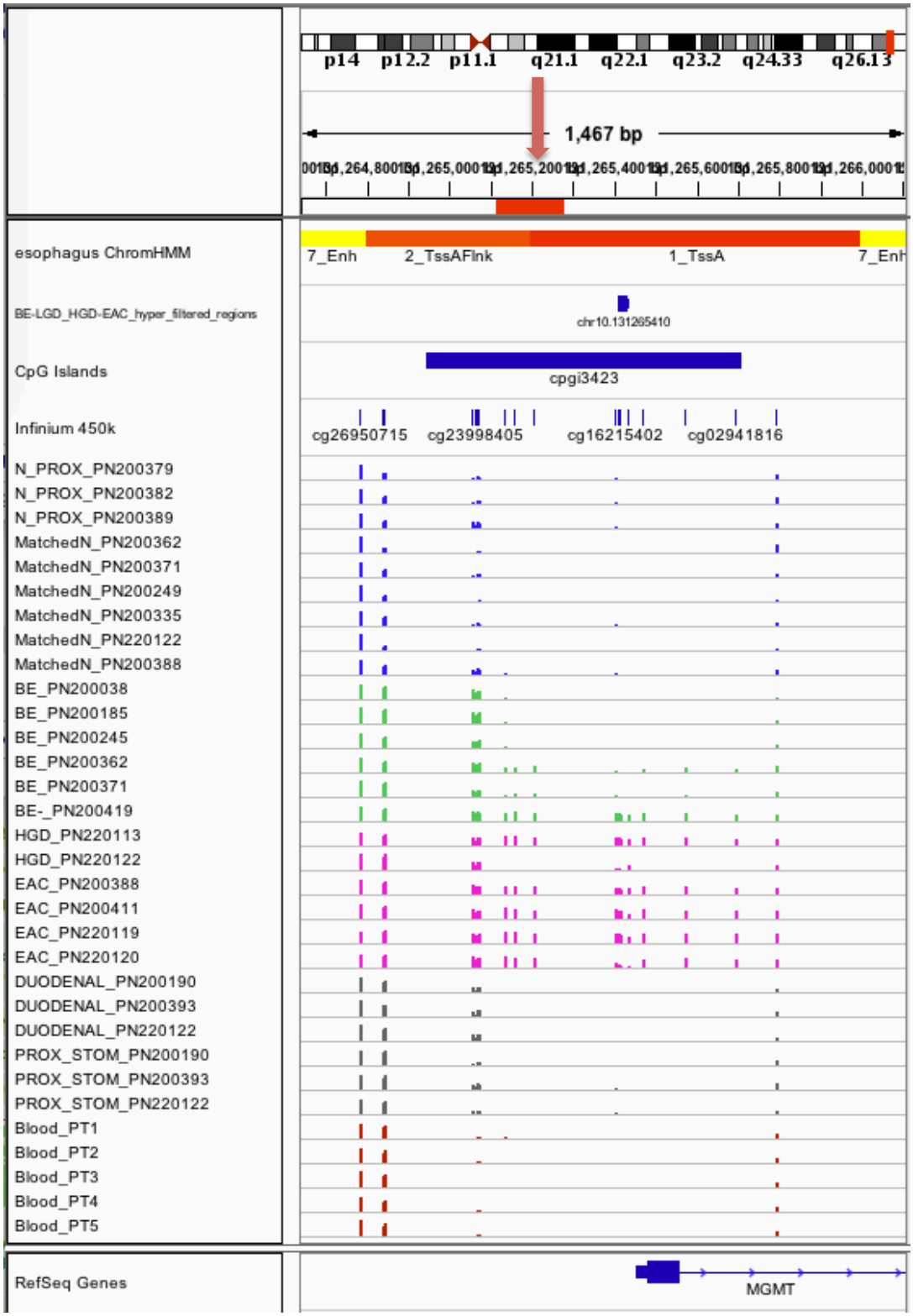


Figure 5-12: Targeted amplicon sequencing of disease-associated differential methylation occurring in the promoter region of O-6-methylguanine-DNA methyltransferase (MGMT). (A) Targeted amplicon sequencing region (seventeen CpG sites, Chr10: 131,265,276 – 131,265,113) showing

independent validation sequencing only (this region was not analyzed for technical validation). Elevated methylation for any form of disease (BE, through dysplastic disease, to EAC) was observed in comparison to the fully unmethylated state of normal, healthy esophageal tissue. (B), (C) This was supported by external validation HM450 data. Elongated validation region examined includes 10 probes, Chr10: 131,265,137 – 131,265,697. Targeted amplicon sequencing region (red arrow), lies over three probes: cg14194875, cg00618725 and cg12434587. (C) The unmethylated status of normal healthy esophageal tissue was also observed for GERD, stomach and normal peripheral blood samples. Heatmap methylation values separated into eight discrete regions 0.000 – 1.000 in increments of 0.125. (D) Genome browser shows HM450 data for a subset of the Training Cohort with >90% tissue sample homogeneity; MGMT amplicon shown in red. Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

5.3.6 Bimodal methylation clustering

Methylation bimodality (hypermethylation of a subset of a particular disease class, with another subset entirely unmethylated) was observed for the majority of target regions analyzed here and supported by genome-wide methylation profiling of a large external validation cohort. Hypermethylated samples across target regions did not correlate with disease severity, nor other clinicopathological factors, such as patient age, gender or presence of *Helicobacter pylori* gastric infection. Thus, it appears that bimodal clustering is target specific and may be due to a variety of reasons. For example, we could hypothesize that the subset of patients with aberrantly methylated VANGL2 may represent a more aggressive dysplasia or adenocarcinoma with worse survival due to reactivation of embryonic development pathways associated with invasion. Similarly, the subset of Barrett's metaplasia, dysplasia and adenocarcinoma patients with promoter MGMT hypermethylation may represent a subset of tissues with greater malignant-change potential due to repressed DNA repair pathways. Furthermore, the inflammation-triggered stem cell origin of Barrett's may result in molecularly

distinct groups of disease arising, which may contribute to the observed bimodal clustering. This is discussed further in Section 4.3.7 Methylation of tumor suppressor p63. We could also hypothesize that chromothriptic events, reported as being more prevalent in EAC than any other cancer (recently reported to be as high as 32%)[90], result in massive genomic rearrangement that could potentially impact methylation machinery in a subset of EAC samples. As a concluding point it is worth noting that this bimodal methylation clustering could also be due to possible 'contamination' of samples with tissue types other than the assigned classification. Despite the stringent cut-off criteria for inclusion of samples in the study, it is possible that the H&E section was not representative of the entire biopsy and thus the molecularly distinct methylomes arising in a sample originate from the different tissue types contributing to the section of biopsy used.

5.3.7 Chromatin state discovery and characterization

ChromHMM was used to determine chromatin state and resultant biological characterization of selected target regions for MiSeq amplicon sequencing (Table 5-12). The majority of selected target regions were located over active transcription start sites (TSS) or flanking active TSS, including the target region ~14Kb upstream of CA4 (Figure 5-13).

Table 5-12: Biological characterization of target regions based on ChromHMM chromatin state discovery. Regions in grey were additional targets included for independent targeted sequencing validation (18 target regions; technical validation cohort evaluated in 10 target regions only). TSS: transcription start site. Target region for MGMT, ISM2 and TNFAIP8L3 across TssA/TssAFlnk boundary. [§]Target region location for CACNA2D2 only 29bp from region defined as active TSS.

	Chr	Start	End	ChromHMM Characterization
N v HGD-EAC				
ELOVL5	6	53,212,967	53,213,085	Active TSS
ZNF790	19	37,329,427	37,329,327	Active TSS
N v Any disease				
SCOC	4	141,294,938	141,294,791	Active TSS
ZNF570	19	37,960,451	37,960,359	Active TSS

	Chr	Start	End	ChromHMM Characterization
MGMT	10	131,265,276	131,265,113	Active TSS / Flanking active TSS
BE v HGD-EAC				
ARL10	5	175,792,604	175,792,424	Active TSS
CACNA2D2	3	50,541,229	50,541,332	Quiescent / Low[§]
ISM2	14	77,965,343	77,965,434	Active TSS / Flanking active TSS
LRRC43	12	122,667,942	122,667,857	Active TSS
TRANK1	3	36,985,944	36,986,065	Active TSS
TUBA3FP	22	21,368,817	21,368,681	Flanking active TSS
KLF7	2	208,031,073	208,031,144	Active TSS
TNFAIP8L3	15	51,385,011	51,385,936	Active TSS / Flanking active TSS
TEPP	16	58,018,911	58,018,827	Flanking active TSS
ZNF699	19	9,420,542	9,420,418	Active TSS
VANGL2	1	160,370,041	160,370,188	Active TSS
ZNF221	19	44,455,289	44,455,448	Active TSS
Upstream CA4	17	58,212,732	58,212,927	Active TSS

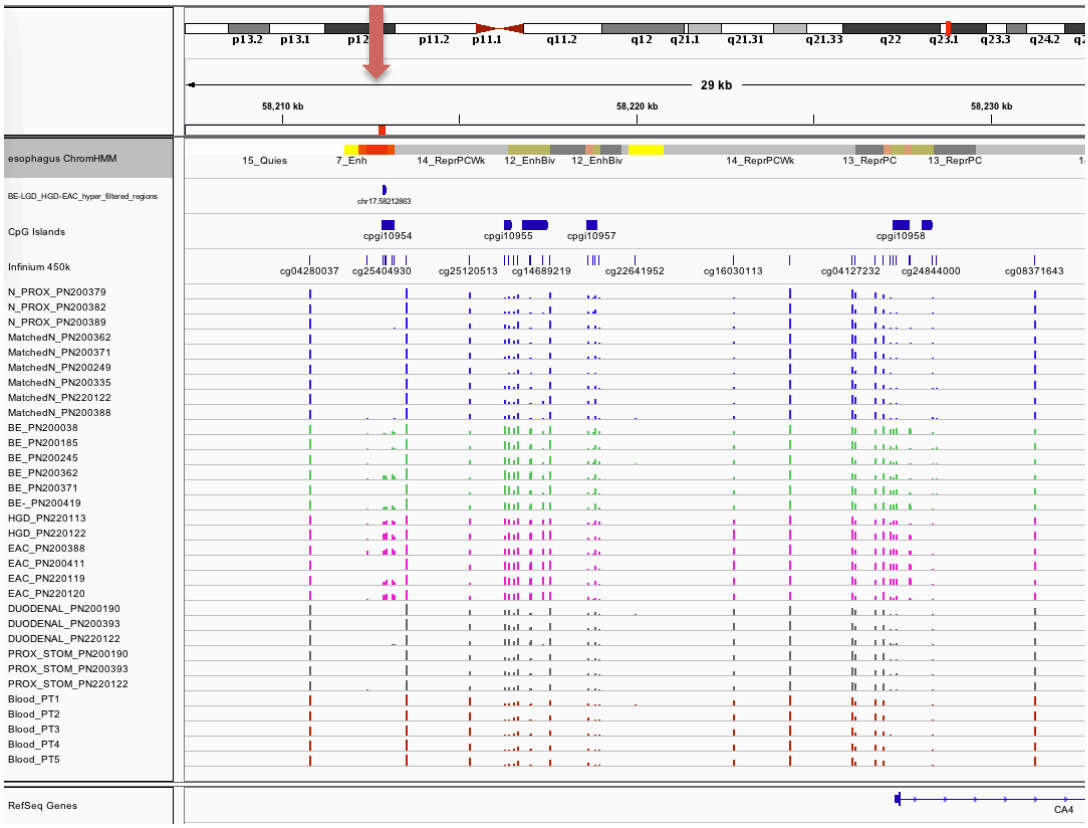


Figure 5-13: Despite being 14kb upstream of carbonic anhydrase IV (CA4), the target region for amplicon sequencing is located within the defined active transcription start site. ChromHMM annotation for normal esophageal mucosa; note that chromatin marks may be disrupted in carcinogenesis and therefore associated regulatory regions may vary. Targeted amplicon

sequencing region Chr17: 58,212,732 – 58,212,927 is marked in red (with arrow). Height of bars indicate percentage methylation (0.0 – 1.0), colour code N: blue, BE: green, HGD and EAC: pink, duodenal and proximal stomach: grey and normal blood: dark red.

5.3.8 Differential expression in selected target regions for validation

Firstly, chromatin state discovery was used to determine the biological characterization of target regions selected for validation (Table 5-12). All regions were across active transcription start sites or flanking active transcription start sites with the exception of CACNA2D2, however the differentially methylated target region for CACNA2D2 was only 29bp from the active transcription start site. Note that ChrommHMM annotation is based on normal esophageal mucosa and chromatin marks may be disrupted by carcinogenesis. We examined gene expression of these regions to examine whether disease-associated differential methylation at, or near the transcription start site resulted in altered expression of these genes. Log fold change and p-values of differential expression of transcript clusters for target region genes are given in Table 5-13 below.

Table 5-13: Differential expression of transcript clusters for target region genes. Target regions and calculated log fold change based on comparison classes as defined in Table 5-5. LogFC: log fold change. Tstat: t-statistic. IT: intronic transcript. AS: anti-sense RNA. Statistically significant differential expression (p≤0.05) in grey.

Transcript Gene	Transcript Cluster ID	LogFC	Tstat	P-value	Transcript ID
ELOVL5	TC06001813.hg.1	-0.245145919	-0.864904524	0.396979192	NM_00124828
ZNF790	TC19001472.hg.1	-0.450496532	-2.767122934	0.011629737	NM_206894
ZNF790-AS1	TC19000507.hg.1	-0.099202052	-1.462030727	0.158711797	NR_040027
SCOC	TC04000689.hg.1	-0.41001175	-1.728479308	0.098753639	NM_001153585
ZNF570	TC19000516.hg.1	-0.106415402	-0.641224589	0.528395537	NM_144694
MGMT	TC10000927.hg.1	-0.010392113	-0.093501137	0.926402275	NM_002412
ARL10	TC05000997.hg.1	-0.03767486	-0.603898245	0.552456155	NM_173664
CACNA2D2	TC03001433.hg.1	-0.185665587	-2.849359038	0.009675928	NM_001005505
ISM2	TC14001347.hg.1	-0.029111072	-0.498198159	0.623583098	NM_182509
LRRC43	TC12000970.hg.1	-0.060884717	-1.099328887	0.28421468	NM_001098519
TRANK1	TC03001283.hg.1	-1.402882517	-4.992270035	6.31E-05	NM_014831
TUBA3FP	TC22000528.hg.1	0.051041094	0.688895904	0.49851898	NR_003608

Transcript Gene	Transcript Cluster ID	LogFC	Tstat	P-value	Transcript ID
KLF7	TC02002705.hg.1	-0.34787389	-1.337845056	0.195423858	NM_003709
KLF7-IT1	TC02002707.hg.1	-0.453495891	-1.638188279	0.116456368	ENST00000428777
TNFAIP8L3	TC15001360.hg.1	-0.06422292	-0.478649142	0.637192575	NM_207381
TEPP	TC16000498.hg.1	-0.006025816	-0.092379042	0.927282866	NM_199046
ZNF699	TC19001145.hg.1	-0.130564682	-0.674957758	0.507153486	BC109268
VANGL2	TC01001369.hg.1	-0.145596905	-1.759018502	0.093311761	NM_020335
ZNF221	TC19000612.hg.1	0.200488844	1.795277556	0.087187766	NM_013359
Upstream CA4	TC17000736.hg.1	-0.12668215	-1.930207555	0.067361484	NM_000717

The majority of differential expression of transcript clusters for target region genes were not significant. The majority of disease-associated hypermethylation, including all statistically significant changes, resulted in a decrease in expression. Given the recent work by Moarii et al[94], these findings are not surprising and support the more complicated nature of the relationship between promoter methylation and gene expression in cancer development.

5.4 Discussion

The results presented in this chapter cover internal validation by targeted amplicon sequencing, using both technical and independent cohorts, as well as external validation by whole genome methylation profiling using a large cohort (250 samples from 154 patients) of the first published HM450 data for esophageal adenocarcinoma mucosa (GSE72874, released 01/04/2016).

Target regions for validation were carefully selected by applying stringent criteria and a series of cut-offs, including minimizing baseline methylation, methylation in normal esophageal mucosa, normal peripheral blood and control tissues and maximizing methylation change. Control tissues ensure selected biomarker targets are disease-associated rather than associated with the maintenance of columnar mucosa. Peripheral blood from normal, healthy individuals ensures applicability of selected targets as blood biomarkers for early disease identification.

Potential esophageal adenocarcinoma methylation biomarkers, as identified by current literature (including p16 (CDKN2A), TMEFFF2 (HPP1), RUNX3, CDH13, TAC1, NELL1, AKAP12, SST, APC, DAPK, CDH1, MGMT, TIMP3,

p57 (CDKN1C), p73, EREG[38, 59, 60, 62, 63]) were examined in our samples to gain more insight into these biomarkers and also evaluate the impact of availability of HM450 technology over the older 27k microarrays. The majority of literature biomarkers exhibited differential methylation with respect to distinguishing 'any disease' from normal esophageal mucosa: Barrett's mucosa was not distinguishable from dysplastic or EAC tissues. Exceptions were CDH1, EREG and CDKN1C which showed no differential methylation in our samples, the possibility of a BE only marker and poor coverage (only 1 EAC sample differentially methylated) respectively. Overall, the majority of these biomarkers were unable to differentiate non-dysplastic BE from dysplasia and EAC, thus ultimately lacking real clinical utility owing to the low rate of carcinogenic progression and benign nature of Barrett's esophagus.

Based on these findings, the HM450 training cohort data set was reanalyzed using the comparison N vs any disease (BE/LGD/HGD/EAC). Strong differential methylation was observed, as expected due to not only disease progression, but also change in tissue mucosal structure. The top 15 differentially methylated regions identified using this comparison all had average baseline methylation (N samples) <0.08 and average delta methylation of >0.60 (maximum 0.69). None of the current literature markers were in the top 15 identified differentially methylated regions. NELL1 was the highest of the literature hits, with average baseline methylation 0.06 and average delta methylation 0.52. The lack of literature hits in the top differentially methylated target regions using this comparison is likely indicative of the much greater coverage provided by HM450 technology.

For comprehensive validation, alternative technology (targeted sequencing rather than the whole genome methylation microarray approach used for discovery) was applied, namely MiSeq amplicon sequencing. The advantage of utilizing a MiSeq targeted sequencing approach is its ability to accurately quantitate the levels of methylated and unmethylated alleles at each CpG site across multiple regions of interest *at single-base resolution*, thus enabling selection of the most clinically appropriate disease-specific biomarkers.

I used unbiased amplification of methylated and unmethylated DNA by bisulfite-specific PCR as a starting point for MiSeq and designed assays to be directly applicable for investigations in blood. By designing short amplicons (72-196bp) and adjusting MiSeq library preparation protocols to suit this, the test is directly applicable for investigating the methylation status of the short, degraded fragments of cell-free circulating DNA extracted from blood plasma.

The majority of targets (13 of 18, 72%) that underwent validation were successfully validated, with technical and independent internal validation as well as large cohort external validation supporting original hypotheses from discovery data. Among these targets are regions with little known regarding function, such as ARL10 and TUBA3FP; and also well-defined regions with known carcinogenic roles, such as zinc finger family member KLF7 and leucine-rich repeat superfamily member LRRC43. Aberrant promoter methylation of many of these regions is novel for EAC carcinogenesis, for example targets such as VANGL2, LRRC43 and KLF7. A developing role for VANGL2 in carcinogenesis has only recently begun to be elucidated, with a role in ER-positive breast cancer uncovered in 2014. LRRC43 has been implicated in a number of different cancers such as colorectal, glioma, melanoma and head and neck cancer, but is novel for esophageal adenocarcinoma. A flurry of publications in 2015 associated various KLF family members with carcinogenic progression in various cancers, including KLF4 in esophageal squamous cell carcinoma and KLF17 in esophageal cancer; however aberrant KLF7 methylation and its role in EAC carcinogenesis is novel.

MiSeq targeted amplicon sequencing enabled elucidation of distinct methylation populations within disease classifications, trending towards either entirely unmethylated or highly methylated. This bimodal methylation was observed for the majority of target regions identified as biomarkers for intervention requiring disease (BE v HGD-EAC). This observation gives rise to hypotheses of specific sub-populations within disease classifications that

may represent more aggressive disease (for example, methylated VANGL2 induces reactivation of embryonic development pathways associated with invasion) or disease with greater malignant potential (for example, methylated MGMT results in repressed DNA repair pathways). An alternative hypothesis for bimodal methylation is the inflammation-triggered stem cell origin of BE. Molecularly distinct groups of disease may arise triggered by reflux-induced inflammation at the gastroesophageal junction, contributing to the observed bimodal methylation clustering observed for BE, dysplastic disease and EAC. One final hypothesis for this bimodal methylation is the high levels of chromothriptic events, reported as being more prevalent in EAC than any other cancer[90]. This chromothripsis may result in massive catastrophic genomic rearrangement that potentially disrupts the methylation machinery in a subset of EAC samples to result in the observed methylation bimodality. Further studies would be required to determine if any, or a combination of these hypotheses, are correct.

Deficient or absent DNA repair is an important underlying cause of cancer[153]. Promoter methylation reducing DNA repair gene expression occurs early in progression to GI cancers[153]. This early hypermethylation is evident in elevated MGMT methylation observed in non-dysplastic Barrett's mucosa and maintained through dysplastic development and progression to adenocarcinoma. It is possible that this DNA repair gene silencing may contribute to the high levels of genomic instability evident in EAC[90]. A future aspect of this work may be in predicting patient response to, and/or suitability for, neoadjuvant chemotherapy. Cancer cells deficient in DNA repair are more vulnerable to inactivation by DNA damaging agents than normal cells, which can be used to clinical advantage. For example, methylation of DNA repair gene MGMT could be investigated as a possible biomarker for predicting patient suitability for neoadjuvant chemotherapy in surgery-requiring esophageal adenocarcinoma. There are many DNA repair pathways that could be used to create a panel of biomarkers to this effect: in 2015, Bernstein et al[153] reported 18 DNA repair proteins, active across 6 different DNA repair pathways, that were subject to epigenetic induced down-regulation in GI cancers.

ChromHMM was used for characterization of validated target regions, the majority (17 of 18, 94%) of which were over active transcription start sites or flanking active transcription start sites. Despite this, very few of these differentially methylated target regions were associated with differentially expressed transcript clusters (3 of 18, 17% with statistically significant down-regulation). Given the recent work by Moarii et al[94], these findings are not surprising and support the more complicated nature of the relationship between promoter methylation and gene expression in cancer development.

This comprehensive validation and characterization of disease-associated differentially methylated target regions shows biomarker viability for early esophageal adenocarcinoma. Through careful consideration of assay design, there is direct transferability of this work for investigation of these target regions in blood.

CHAPTER 6: METHYLATION BIOMARKERS FOR ESOPHAGEAL ADENOCARCINOMA AND PRECURSOR DISEASE

6.1 Introduction

As outlined in Chapter 1 there is an urgent need for a better way to diagnose early stage esophageal adenocarcinoma. Current endoscopic/histologic methods are not viable for high-risk population screening, resulting in late stage diagnosis, contributing to the poor outcome for this disease.

Carcinogenic progression for esophageal adenocarcinoma is not straightforward. High population prevalence of non-dysplastic precursor disease, but low progression rates to dysplasia are complicating factors for a screening test. Identification of asymptomatic, benign BE is not necessary if a reliable, robust method for low and high-grade dysplasia identification exists. The most clinically relevant biomarkers would differentiate patients with dysplastic Barrett's (low and high-grade) as well as EAC from those non-dysplastic Barrett's or a normal, healthy esophagus. This is the focus of my thesis.

6.1.1 Diagnostic EAC biomarkers

The current state of play for early diagnostic biomarkers for EAC is nicely summarized by Shah et al in a 2013 review[202], applying the National Cancer Institute's Early Detection Research Network (EDRN) guidelines for biomarker discovery and development (a 5-phase process) to show development status of EAC diagnostic biomarkers (Figure 6-1). In brief, Phase I includes preclinical exploratory studies (normal versus tumor analyses), Phase II: clinical assay development and validation, Phase III: retrospective longitudinal repository studies (assay applied to prospectively collected stored samples to assess biomarker utility), Phase IV: prospective screening and Phase V: cancer control studies (large-scale clinical trials).

Using these guidelines, this study falls into Phases I and II of the EDRN biomarker development process. The value in reverting essentially 'back to the start' for this study lies in its focus on inclusion of dysplastic BE samples for profiling and changing the current 'tumor-normal' focus to the more clinically appropriate 'intervention requiring disease vs non-dysplastic, benign BE' focus. Implementing HM450 profiling of LGD and HGD is novel for early EAC detection biomarker studies and may be key to identification of biomarkers with true clinical utility that retain robustness and sensitivity/specificity as they progress towards clinical implementation.

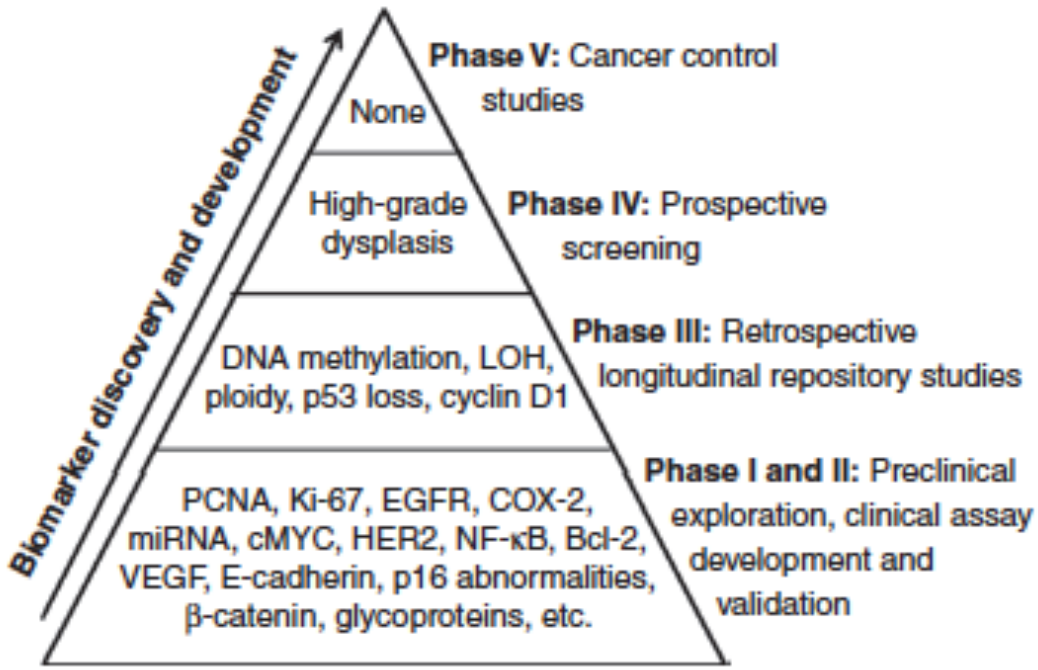


Figure 6-1: Summary of current diagnostic biomarkers for BE and EAC with respect to the Early Detection Research Network (EDRN) clinical phases of development. Figure courtesy of Shah et al[202] (2013) review: 'Early diagnostic biomarkers for esophageal adenocarcinoma – the current state of play'. For early diagnostic EAC biomarkers, DNA methylation is one of the better performing biomarkers, and has progressed as far as Phase III studies.

6.1.2 Prediction biomarkers in EAC

Possibly the most progressed of all EAC biomarkers are those for predicting risk of progression from BE to EAC. As outlined in Chapter 1, a panel of eight DNA methylation biomarkers to predict the risk of progression from BE to EAC, proposed by Stephen Meltzer and his team at John Hopkins University[38] has been launched commercially in 2013. A significant limitation of the Meltzer laboratory test is its reliance on endoscopic tissue biopsies. The test is intended be used as a supplement to current endoscopic/histology diagnosis and surveillance; and may reduce load on heath care systems by increasing surveillance interval of low-risk progression patients. However, the impact on outcome for EAC sufferers has not been evaluated and may not be significant as the test does not address disease detection in a high-risk population, but rather focuses on patients with *known* non-dysplastic disease (who are likely already in surveillance programs and would have any progressive disease detected by standard endoscopic/histologic surveillance at an early, curable stage).

A more informative prediction biomarker for EAC may be surrounding patients diagnosed with LGD. There is disagreement in guidelines worldwide regarding treatment of LGD, ranging from a surveillance-only approach to an aggressive treatment approach (removal of dysplastic tissue by endoscopic mucosal resection (EMR) or radio-frequency ablation (RFA))[52, 81]. A biomarker able to predict the necessity of treatment for LGD, for example if LGD would naturally regress to non-dysplastic disease without treatment versus LGD likely to progress to HGD and EAC, would be valuable in guiding treatment decision and result in better use of specialist time. There is no necessity for a biomarker such as this to be blood-based; LGD patients would be undergoing endoscopic surveillance and tissue biopsies taken routinely, thus a tissue-based biomarker is acceptable in this situation. A biomarker for determination of treatment regime for LGD patients would be used as an adjunct to current surveillance and used to guide subsequent surveillance or treatment to avoid progression to EAC.

Furthermore, biomarkers for monitoring disease status during treatment would also be clinically valuable in the current EAC climate. A biomarker for this purpose may be used to assess whether current treatment regimes are effective or if a more aggressive treatment approach is required. A blood-based biomarker here would be valuable, however EMR and RFA treatment require endoscopic examination and thus even a tissue-based biomarker may have some level of clinical utility, used as an adjunct to current endoscopic/histologic disease surveillance methods.

This chapter aims to not only evaluate methylation biomarker performance with regard to diagnostic utility (investigating biomarkers individually and multiplexed into panels); but also investigate applications such as those described above by performing full clinical follow up on all patients used in the study to determine possible application for disease progression prediction (non-dysplastic to dysplastic BE as well as LGD progression or natural regression) as well as possible application in disease monitoring with treatment.

6.1.3 Chapter 6 aims

To evaluate DNA methylation biomarker performance with regard to diagnostic utility and application for disease progression prediction and disease monitoring with treatment.

6.2 Methods

6.2.1 Probe level biomarker filtering

In addition to a global analysis of genome-wide methylation profiling data, more specific biomarker-focussed analysis was performed. Biomarker filtering cut-offs (hypermethylation) were applied to probes from each comparison as outlined in Table 6-1. Baseline methylation (baseline groups: N or BE as per comparison group) cut-offs were applied, then differentially methylated regions (DMRs) were ranked and filtered to remove those with methylation evident in normal peripheral blood (using the publicly available

data set GEO accession GSE48472)[110], with foresight for a non-invasive blood biomarker test. Differentially methylated probes were further filtered to remove those with methylation above baseline in control (duodenal and proximal stomach) tissues, as well as normal squamous epithelium (for BE v HGD-EAC comparison). Integrative Genomics Viewer (IGV, Broad Institute, Cambridge, USA) was used to visualise and explore HM450 methylation data[203]. No exclusion criteria regarding differential methylation in neighbouring probes was applied (as it was for selection of regions for validation by targeted sequencing, Section 5.2.1 Target region selection).

Table 6-1: Cut-off criteria (hypermethylation) applied for determination of clinically relevant biomarkers. Cut-offs apply to average methylation values detected for each probe within the defined target region.

	N v HGD-EAC	**N v BE**	**BE v HGD-EAC**
Baseline methylation (β)	≤ 0.10	≤ 0.10	≤ 0.10
Methylation change (Δβ)	≥ 0.50	≥ 0.50	≥ 0.20
Normal blood (β)	≤ 0.10	≤ 0.10	≤ 0.10
Normal esophageal tissue (β)	≤ 0.10	≤ 0.10	≤ 0.10
Duodenal tissue (β)	≤ 0.10	≤ 0.10	≤ 0.10
Proximal stomach tissue (β)	≤ 0.10	≤ 0.10	≤ 0.10

6.2.2 Determination of biomarker performance

Sensitivity and specificity for each of the individual target regions was calculated based on MiSeq targeted sequencing data for methylation at individual CpG sites within each region. Average methylation at each probe was compared to a predefined cut-off (Table 6-2) and assigned a true negative (TN, Avg β<specified cut-off) or false positive (FP, Avg β>specified cut-off) for baseline samples or a true positive (TP, Avg β>specified cut-off) or false negative (FN, Avg β<specified cut-off) for disease samples. A positive result was recorded when ANY probe within the specified region returned an average methylation above the designated cut-off for that region. Cut-offs were assigned to achieve maximum sensitivity with minimum impact on specificity. The performance of individual target regions was evaluated by determining sensitivity (percentage of TP with respect to total disease

samples) and specificity (percentage of FP with respect to total baseline samples). The performance of various multiplexed target regions was evaluated using sensitivity and specificity for individual target regions; regions were selected based on attaining maximum sensitivity with minimum impact on specificity. Cut-offs, TN, FP, TP and FN counts were as for individual target regions.

6.2.3 Evaluation of biomarkers for prediction of progression and disease monitoring

Clinical follow up was performed on all patients who contributed samples to either technical or independent validation cohorts and had associated MiSeq targeted sequencing data (technical validation cohort patients also had HM450 and HTA2.0 profiling data). Average methylation at each CpG site of patients with changed clinical diagnoses were compared to averages of all other patients with those particular diagnoses. Unpaired t-tests or ordinary one-way ANOVA were performed to evaluate if aberrant methylation was statistically significant.

6.3 Results

6.3.1 Biomarker determination

Heat maps were compiled for all disease-associated hypermethylated probes meeting specified cut-off criteria (Table 6-1) for each of the comparisons: N v HGD-EAC (Figure 6-2, 799 probes), N v BE (Figure 6-3, 235 probes) and BE v HGD-EAC (Figure 6-4, 165 probes), based on genome-wide methylation profiling data of the training cohort. A Venn diagram (Figure 6-5) shows the overlap in identified biomarker probes between each of the comparison groups. Details of all hypermethylated biomarker probes meeting cut-off criteria for each comparison are given in Appendix 7: Hypermethylation biomarker probes for BE and EAC. GENCODE v19 was used for reference human genome annoation[204].

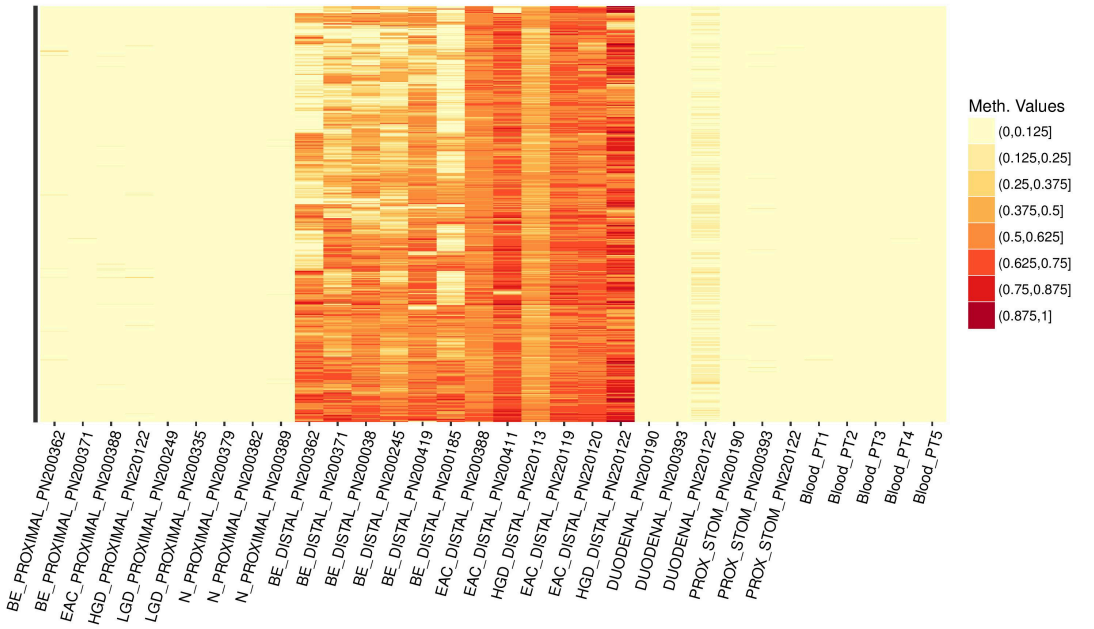


Figure 6-2: Hypermethylated biomarker probes (n=799) for differentiation of HGD and EAC from N ($\Delta\beta \geq 0.50$). Baseline restrictions applied to normal tissue (unmethylated, $\beta \leq 0.05$) as well as control (duodenal and proximal stomach) tissues and blood from normal healthy patients ($\beta \leq 0.10$). No restriction placed on methylation level in BE patients. Samples shown are a subset of the training cohort with >90% tissue sample homogeneity. All samples taken proximally are normal, healthy esophageal mucosa. Methylation values separated into eight discrete regions 0.000 – 1.000 in increments of 0.125.

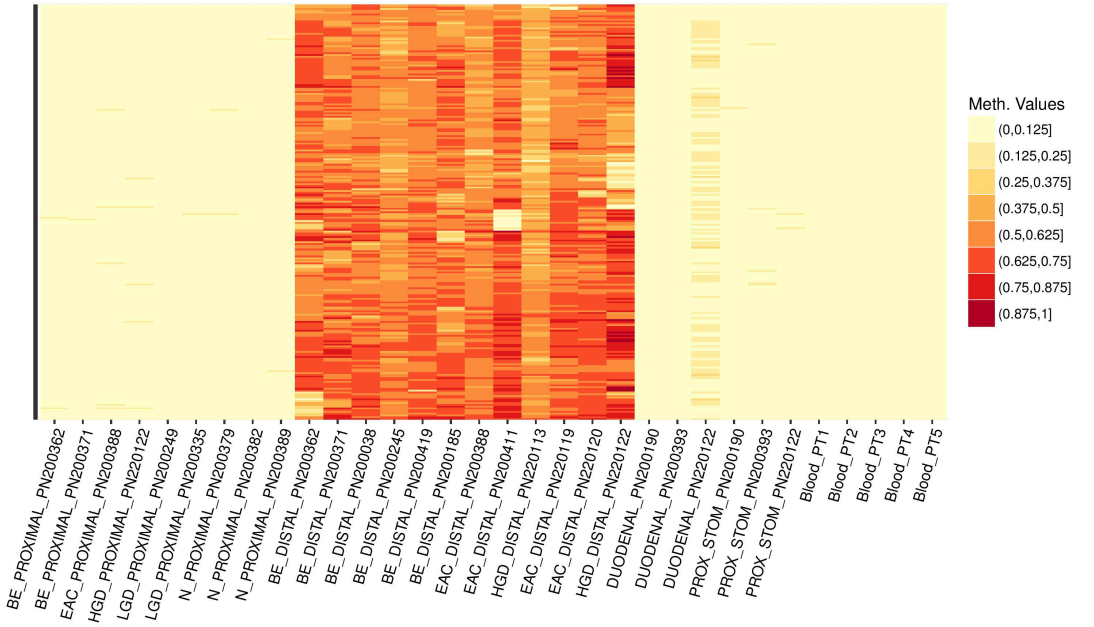


Figure 6-3: Hypermethylated biomarker probes (n=235) for differentiation of BE from N ($\Delta\beta \geq 0.50$). Baseline restrictions applied to normal tissue (unmethylated, $\beta \leq 0.05$) as well as control (duodenal and proximal stomach) tissues and blood from normal healthy patients ($\beta \leq 0.10$). No restriction placed on methylation level in HGD and EAC patients. Samples shown are a subset of the training cohort with >90% tissue sample homogeneity. All samples taken proximally are normal, healthy esophageal mucosa. Methylation values separated into eight discrete regions 0.000 – 1.000 in increments of 0.125.

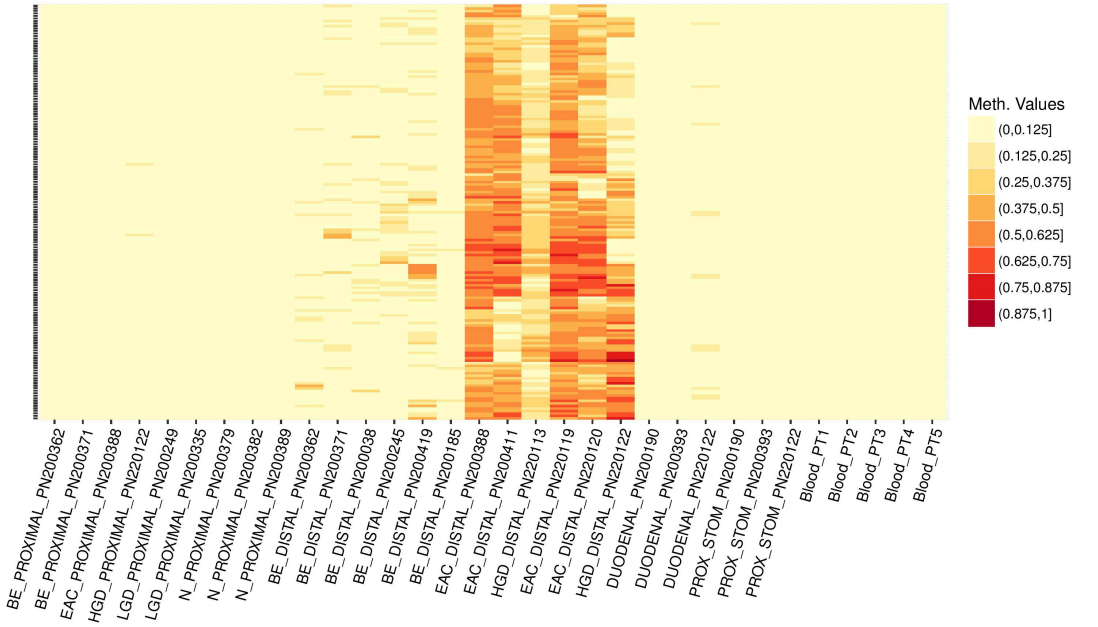


Figure 6-4: Hypermethylated biomarker probes (n=165) for differentiation of intervention requiring disease (HGD and EAC) from BE and N ($\Delta\beta \geq 0.20$). Baseline restrictions applied to non-dysplastic BE tissue (unmethylated, $\beta \leq 0.10$) as well as control (normal esophageal epithelium, duodenal and proximal stomach) tissues and blood from normal healthy patients ($\beta \leq 0.10$). Samples shown are a subset of the training cohort with >90% tissue sample homogeneity. All samples taken proximally are normal, healthy esophageal mucosa. Methylation values separated into eight discrete regions 0.000 – 1.000 in increments of 0.125.

Considerable overlap (n=168) was present in hypermethylation biomarker probes differentiating BE and HGD-EAC from normal healthy esophageal tissue (Figure 6-5). 71.49% of hypermethylated probes for BE identification also identify HGD and EAC, meaning the majority of disease-associated aberrant methylation occurring early in the metaplasia-dysplasia-carcinoma sequence is maintained. Conversely, only 21.03% of hypermethylated probes for HGD-EAC identification also identify BE, indicating that much of the significant aberrant methylation associated with high-grade dysplasia and esophageal adenocarcinoma is developed late in the metaplasia-dysplasia-carcinoma sequence. As can be seen in Figure 6-2, many of the HGD-EAC biomarkers show some level of increased methylation in BE, however due to

hypermethylation criteria Δβ ≥ 0.50, these biomarkers are not identified as being associated with BE. Thus, it appears that methylation becomes more prolific in these regions as disease progresses.

The majority of clinically relevant hypermethylation probes for the detection of intervention requiring disease (differentiating HGD and EAC from non-dysplastic BE and normal healthy esophageal tissue) are unique to this comparison (91.5%, 151 of 165 unique). Forcing baseline methylation of N and non-dysplastic BE to be ≤ 0.10 results in lower average Δβ methylation to HGD-EAC samples. An overlap of 14 differentially methylated probes with the N v HGD-EAC comparison was observed, indicating higher levels of methylation in HGD and EAC for these particular probes (≥ 0.50). Note that region selection for validation by targeted amplicon sequencing included criteria pertaining to identification of aberrant methylation across groups of probes (within 300bp, at least 2 differentially methylated probes required) as well as region suitability for BSP assay design. Hence some of the top aberrantly hypermethylated individual probes listed in Appendix 7: Hypermethylation biomarker probes for BE and EAC were not part of regions selected for validation.

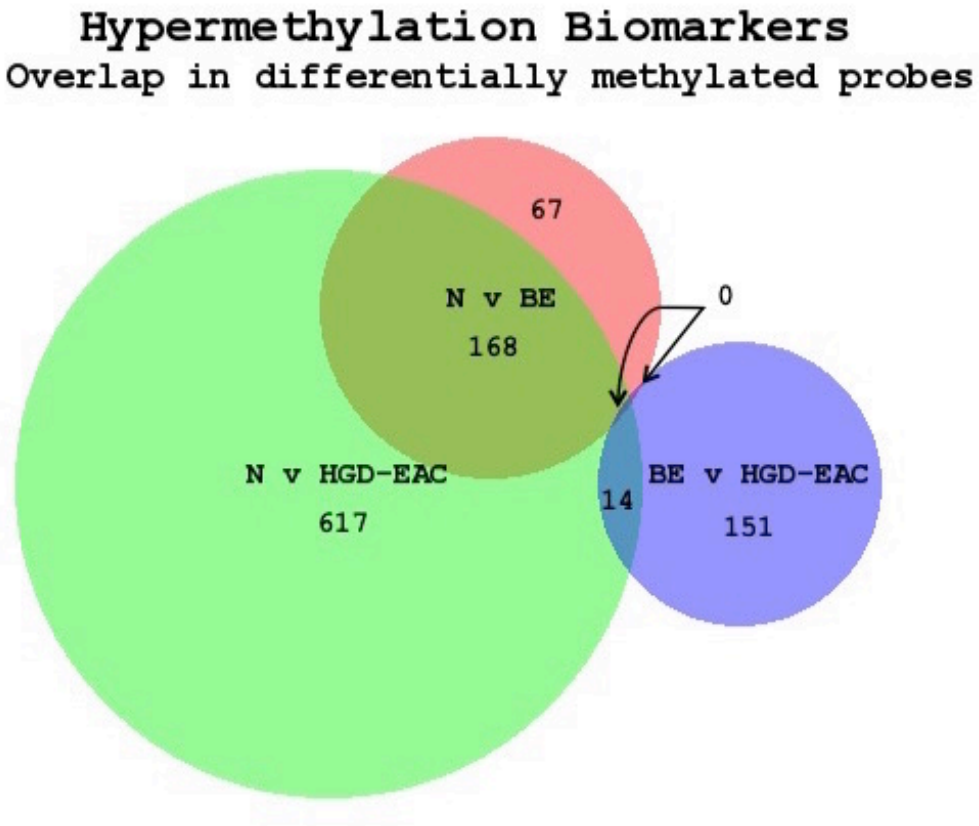


Figure 6-5: Venn diagram showing overlap in identified hypermethylated biomarker probes between each of the comparison groups. Totals: N v BE n=235 probes, N v HGD-EAC n=799 probes, BE v HGD-EAC n=165 probes. Considerable overlap was present in probes for identification of BE and HGD-EAC from N, sharing 168 common regions, whereas the majority of clinically relevant probes for differentiating intervention requiring disease from normal and non-dysplastic disease were unique to the comparison. Venn diagram created using BioVenn[205].

6.3.2 Performance of individual target regions

Sensitivity and specificity of each individual target region (and the cut-off used) are given in Table 6-2. Analysis was performed both with and without LGD patients (included with HGD and EAC as intervention requiring disease) to determine whether target regions were capable of detecting these difficult to classify patient samples. Sensitivity and specificity data for MGMT has been determined using the 'N v Any disease' comparison due to observed elevated methylation of non-dysplastic BE, confirming results reported in

literature[93, 151]. Results for VANGL2 exclude the CpG site at Chr1: 160,370,070 due to assay positioning restrictions (unable to design assay to region of interest due to ubiquitous poly-C repeats, assay moved to CpGi1843 shore).

The performance of individual target regions as stand-alone biomarkers for detection of HGD-EAC or 'Any disease' (non-dysplastic BE through to EAC) is relatively high with reported sensitivity and specificity of ≥ 80% for the five regions. However the majority of these (with the exception of MGMT, β cut-off 0.02) rely on arbitrary methylation cut-offs 0.15 – 0.45; meaning that low level methylation exists in normal squamous esophageal tissue. MGMT methylation is a reliable marker (94.1% sensitivity, 80.0% specificity) for evidence of any disease (non-dysplastic as well as dysplastic BE and EAC) in tissue biopsies. Methylation beta-value cut-off of 0.02 for this assay indicates the region is unmethylated in normal healthy esophagus, but become methylated with metaplastic change and maintains methylation throughout dysplastic progression and adenocarcinoma development.

The performance of individual target regions as stand-alone biomarkers for differentiation of intervention requiring disease from normal and non-dysplastic disease is more complicated. Reported sensitivities (42.9 – 92.3%) reflect the bimodal methylation observed in targeted sequencing results (easily visualized in Figure 5-9(D)). These biomarkers have been shown to be either highly or un-methylated; a trend seen across all 13 regions and discussed in Chapter 5: Target Region Validation and Characterization. However these shortcomings are overcome by multiplexing of target regions, with just three biomarkers needed to identify 100% of HGD and EAC in our technical and independent validation cohorts (refer to Section 6.3.3 Performance of multiplexed target regions).

Table 6-2: Sensitivity and specificity of individual target region methylation for disease detection. Data pertains to the listed comparison classes only. LGD sensitivity / specificity include LGD alongside HGD and EAC for identification

of any dysplastic disease and adenocarcinoma. Marked regions (*) have results from the independent cohort only; all other regions have data from both the technical and independent cohorts. Data sorted by highest sensitivity within comparison group.

	Cut-off	Sens (%)	True positive count	Total disease samples	Spec (%)	False positive count	Total baseline samples	LGD Sens (%)	LGD Spec (%)
N v HGD-EAC									
ZNF790	0.20	100.0	13	13	85.7	2	14	100.0	85.7
ELOVL5	0.15	84.6	11	13	100.0	0	14	75.0	100.0
N v Any disease									
MGMT*	0.02	94.1	16	17	80.0	1	5		
SCOC	0.45	90.6	29	32	100.0	0	14		
ZNF570	0.40	87.5	28	32	85.7	2	14		
N-BE v HGD-EAC									
ISM2	0.10	92.3	12	13	57.7	11	26	85.0	57.7
TUBA3FP	0.05	84.6	11	13	84.6	4	26	80.0	84.6
LRRC43	0.08	84.6	11	13	80.8	5	26	65.0	80.8
VANGL2*	0.02	71.4	5	7	100.0	0	11	63.6	100.0
ARL10	0.08	69.2	9	13	100.0	0	26	55.0	100.0
CACNA2D2	0.08	61.5	8	13	84.6	4	26	45.0	84.6
TRANK1	0.25	61.5	8	13	73.1	7	26	55.0	73.1
KLF7*	0.04	57.1	4	7	90.9	1	11	45.5	90.9
ZNF699*	0.015	57.1	4	7	81.8	2	11	54.5	81.8
ZNF221*	0.03	57.1	4	7	72.7	3	11	63.6	72.7
Upstream CA4*	0.07	42.9	3	7	90.9	1	11	36.4	90.9
TNFAIP8L3*	0.30	42.9	3	7	90.9	1	11	27.3	90.9
TEPP*	0.04	42.9	3	7	72.7	3	11	54.5	72.7

6.3.3 Performance of multiplexed target regions

Sensitivity and specificity was also considered for a ‘multiplex’ of target regions, selecting panels based on attaining maximum sensitivity with minimum impact on specificity. Two panels were selected, one for detection of intervention requiring disease (differentiating HGD and EAC from N and BE), the other for detection of any dysplastic disease or adenocarcinoma (differentiating LGD, HGD and EAC from N and BE), results in Table 6-3. As previously, VANGL2 results exclude Chr1: 160,370,070 CpG methylation due

to assay positioning restrictions (unable to design assay to region of interest due to ubiquitous poly-C repeats, assay moved to CpGi1843 shore).

Multiplexing target regions significantly improves ability to accurately detect disease-associated differential methylation, overcoming problematic bimodal methylation observed for individual targets. A triplex of hypermethylation in any of TUBA3FP, VANGL2 or ARL10 (regions as per Table 5-5) was able to accurately identify all HGD and EAC patients with 100% sensitivity (beta value cut-offs of 0.05, 0.02 and 0.08 respectively). Test specificity of 84.6% was reported (based on methylation detected in patients with a normal healthy esophagus, or non-dysplastic Barrett's). For the detection of any dysplastic disease or adenocarcinoma, the panel was adjusted and hypermethylation above cut-off in any of TUBA3FP, ARL10 and ZNF699 (beta-value cut-offs 0.05, 0.08 and 0.015 respectively) was used. This triplex was successfully able to identify 95% of all LGD, HGD and EAC patients, with 76.9% specificity (as previously, based on methylation detected in patients with a normal healthy esophagus or non-dysplastic Barrett's). A single LGD patient (PN200335, variable homogeneity reported, as low as 5-10% LGD) was not methylated above designated cut-off in any of the target regions, therefore 95% sensitivity (19 of 20 samples) is maximal for this comparison.

Table 6-3: Sensitivity and specificity of multiplexed target region methylation for disease detection. Triplex was required for both comparisons to achieve maximum sensitivity. A single LGD patient was not methylated above cut-off levels in any of the target regions tested, hence 95% sensitivity is maximal for this comparison. Cut-offs, TN, FP, TP and FN counts as designated for individual regions (Table 6-2).

	Sens (%)	**True positive count**	**Total disease samples**	**Spec (%)**	**False positive count**	**Total baseline samples**
Multiplex for detection of intervention requiring disease (N-BE v HGD-EAC)						
TUBA3FP + VANGL2 + ARL10	100.0	13	13	84.6	4	26

	Sens (%)	True positive count	Total disease samples	Spec (%)	False positive count	Total baseline samples
Multiplex for detection of dysplastic disease and EAC (N-BE v LGD-HGD-EAC)						
TUBA3FP + ZNF699 + ARL10	95.0	19	20	76.9	6	26

6.3.4 Methylation biomarkers for prediction of non-dysplastic to dysplastic disease progression

There is evidence of application of these biomarkers for predicting progression of non-dysplastic to dysplastic BE. Clinical follow up on all patients that contributed a non-dysplastic BE sample to either the technical or independent validation cohorts revealed a single patient who progressed to dysplastic BE, approximately 1 year later (Figure 6-6).

The patient, male, aged 60, had no known family history of BE or EAC, did not suffer from reflux or heartburn at the time the sample was taken (Nissen fundoplication at 38 years of age), and consented to the study for the first time at a pre-operative 'check endoscopy' prior to sleeve gastrectomy (weight loss surgery) at which time his BMI was 55.2kg/m^2 (181cm, 181kg), explaining the high frequency of subsequent visits. The decision to perform a sleeve gastrectomy was difficult as this form of intervention is known increase reflux and worsen BE, however due to his BMI, hypertension, Type 2 diabetes mellitus, obstructive sleep apnea requiring continuous positive airway pressure therapy and other indications, weight loss intervention was prioritized. At his first consenting visit, BE (C4 M5) with no evidence of dysplasia or malignancy was identified.

At visit 2 (endoscopy performed due to vomiting and eating difficulties), the procedure was abandoned early due to complications and no research biopsies were taken. The patient had lost 20kg and was diagnosed with Grade A reflux esophagitis and BE (C3 M3) (endoscopic classification only, histological classification not possible (no biopsies taken)). At visit 3, BE (C1

M5) was identified and LGD at 40cm was confirmed by hospital histopathology (Figure 6-6).

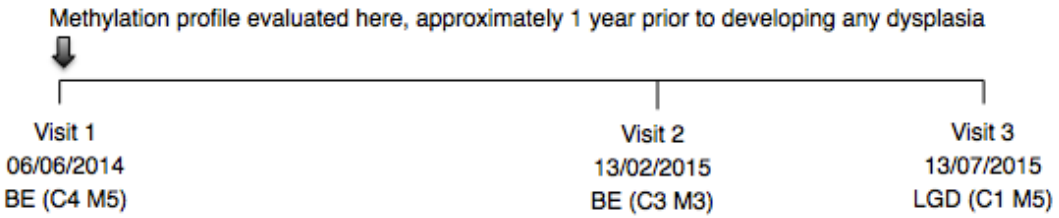


Figure 6-6: Patient timeline for non-dysplastic to dysplastic disease progression. The sample contributing to our study was taken at visit one, when the patient was diagnosed with non-dysplastic BE. Patient classification at visit 3 based on histopathology from hospital biopsies.

The methylation profile of testis, prostate and placenta expressed (TEPP) and target region upstream of carbonic anhydrase IV (Upstream CA4) showed significant increase ($p<0.0001$, both regions) in methylation of non-dysplastic BE, taken one year *before* development of low-grade dysplasia (compared to all other non-dysplastic BE samples where patients did not progress over the course of this study). Progression patient separated and labeled 'BE pro' in Figure 6-7.

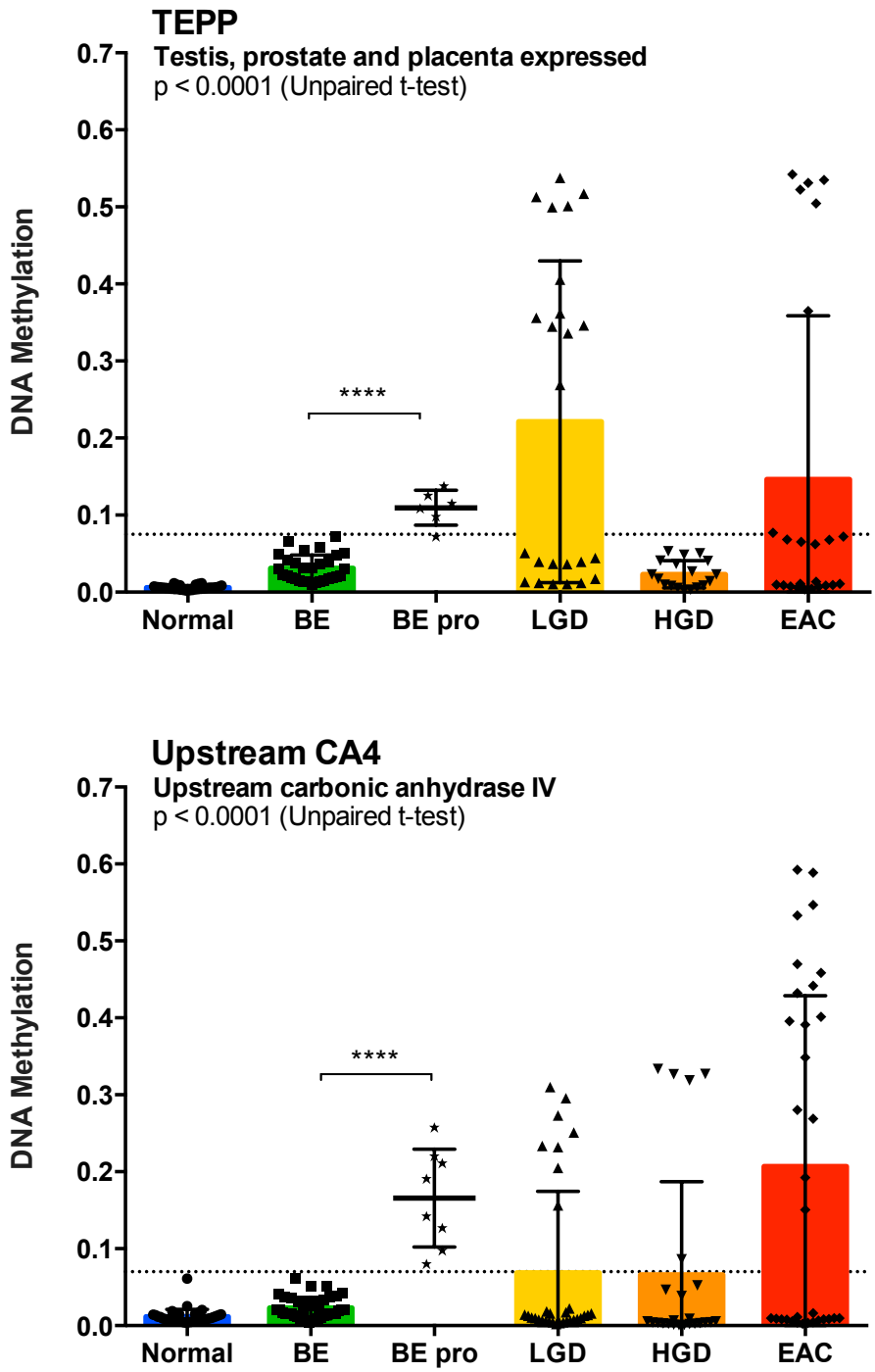


Figure 6-7: Methylation biomarker target regions TEPP and Upstream CA4 for predicting progression from non-dysplastic to dysplastic disease. Average methylation at each CpG site within the target region (TEPP 6 sites, upstream CA4 8 sites) plotted for all patients in the independent validation cohort (24 samples from 15 patients). Statistical significance was evaluated

using an unpaired t-test. BEpro: results from non-dysplastic BE patient with dysplastic progression 1 year subsequent to analysis of this sample.

These results provide evidence that aberrant methylation of these target regions may indicate predisposition to dysplastic progression, of particular clinical significance in a case such as the one described here. This patient has been flagged for follow up: DNA has been extracted from a tissue sample taken at visit 3 (LGD diagnosis) and sent to collaborators in Queensland (A/Prof. Andrew Barbour and colleagues) for HM450 profiling. This data has been included on a collaborative NHMRC grant application to investigate the potential of these biomarkers for disease prediction.

6.3.5 Methylation biomarkers for prediction of the necessity of treatment for low grade dysplasia

The ability to predict progression from low to high grade dysplasia is of particular clinical importance as it may aid in determining how to treat a patient with a LGD classification (still disputed and discussed in guidelines for the management of Barrett's worldwide)[52, 81]. Similarly, biomarkers for the natural regression of LGD to non-dysplastic Barrett's are of clinical interest as these markers, combined with markers for progression from LGD to HGD, could be used together to determine treatment and/or surveillance regime for patients of this controversial classification.

6.3.5.1 Low grade to high grade progression

There is evidence of application of these biomarkers for prediction of progression from low to high-grade dysplasia. Clinical follow up on all patients who contributed an LGD sample to either the technical or independent validation cohorts revealed a single patient who progressed from low- to high-grade dysplasia during the course of this study (Figure 6-8).

The patient, male, aged 66, had no known family history of BE or EAC and consented to the study for the first time when he was being treated for his long-segment LGD (C11 M11). At this, and all three subsequent visits, he

was treated by endoscopic mucosal resection. Focal HGD was first identified approximately 8 months after his first visit (in one of six samples). This was confirmed approximately 3 months later, with focal high grade dysplasia observed in one of eight fragments sent for hospital histopathology (extensive low-grade dysplasia also still present).

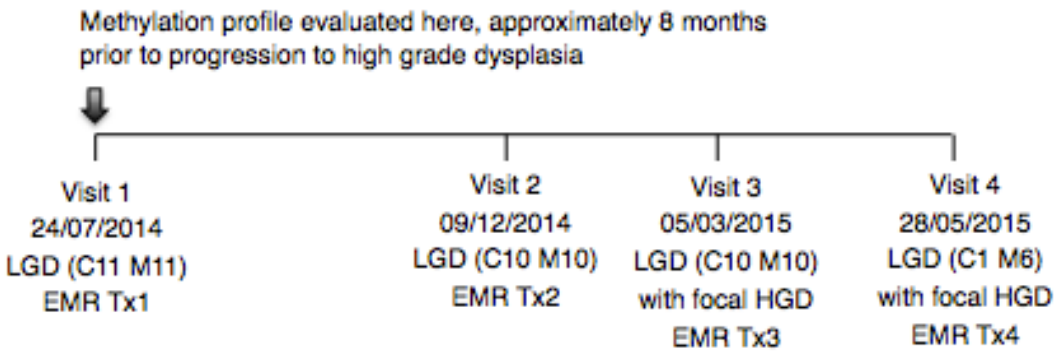


Figure 6-8: Patient timeline for low to high-grade dysplastic progression. The sample contributing to our study was taken at visit one, when the patient was diagnosed with LGD. Patient classification at subsequent visits based on histopathology from hospital biopsies. Focal HGD was identified and treated at visit 3, but persisted at visit 4 (where further treatment was also performed). EMR: Endoscopic mucosal resection, Tx: treatment.

Increased methylation of target regions in leucine rich repeat containing 43 (LRRC43) and zinc finger protein 699 (ZNF699) for this LGD patient are consistent with the profile of more advanced disease. The patient progressed to HGD approximately 8 months later, despite several endoscopic mucosal resection treatments for his long segment LGD (Figure 6-9).

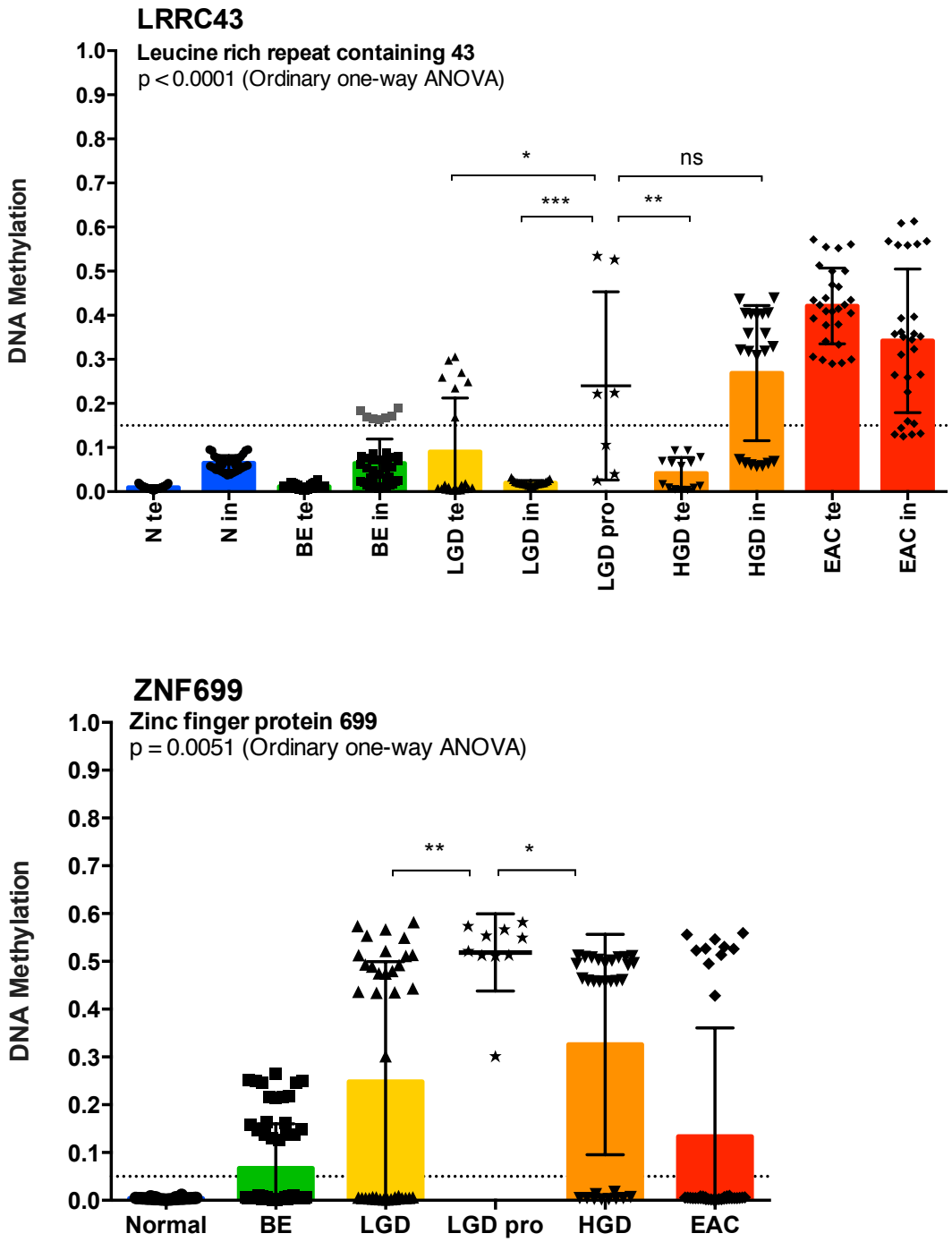


Figure 6-9: Methylation biomarker target regions LRRC43 and ZNF699 for predicting progression from low to high-grade dysplasia. Average methylation at each CpG site within the target region (LRRC43 7 sites, ZNF699 10 sites) plotted for all patients in the technical (te) and independent (in) validation cohorts (LRRC43: total of 48 samples from 33 patients) and the independent cohort only (ZNF699: 24 samples from 15 patients). Statistical significance evaluated using an ordinary one-way ANOVA test and Dunnett's multiple comparisons test, with a single pooled variance. Grey data points (LRRC43)

pertain to a patient that regressed from HGD to LGD to non-dysplastic BE with treatment (refer to Section 6.3.6 Methylation biomarkers for disease monitoring with treatment), indicating that the global methylation profile of this patient has not quite returned to that of non-dysplastic BE, showing remnants of previous higher level disease.

These results provide evidence that increased methylation of these target regions in LGD patients may indicate predisposition to disease progression, thus warranting a more aggressive treatment approach. It is also plausible that these samples contain different proportions of cell sub-types with high and low methylation at these target sites. A full study over an elongated time period with increased numbers of patients with progressive disease is needed to confirm this hypothesis.

6.3.5.2 Natural regression from low grade to non-dysplastic disease

There is evidence of application of these biomarkers for prediction of natural regression from low-grade dysplasia to non-dysplastic disease. Clinical follow up on all patients that contributed an LGD sample to either the technical or independent validation cohorts revealed a single patient who regressed from LGD to non-dysplastic BE (without treatment) during the course of this study (Figure 6-10).

The patient, male, aged 56, had no known family history of BE or EAC and consented to the study for the first time when he was undergoing surveillance for his long-segment LGD (C8 M8). Notable was a small ulcer (biopsied, non-malignant) and bleeding Barrett's. At this, and three subsequent visits, his LGD was extensively biopsied but not treated. His Barrett's fluctuated in circumferential and maximal extent, with extensive LGD identified (v1: 30-34cm, v2: 30-43cm, v3: 31-35cm, v4: 32-39cm). At visit 5, no dysplasia was observed (biopsies from 32-38cm). The patient has a follow up endoscopy scheduled for March 2016.

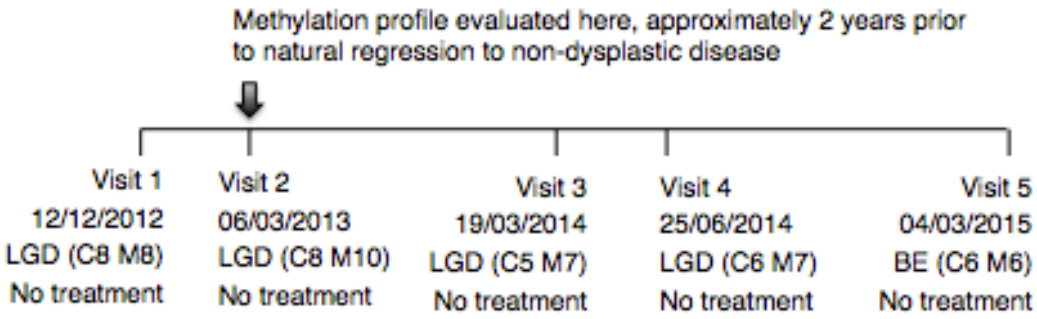


Figure 6-10: Patient timeline for natural regression from low grade to non-dysplastic Barrett's. The sample contributing to our study was taken at visit 2; the patient was suffering from persistent LGD but was not undergoing any form of treatment. Patient classification at preceding and subsequent visits based on histopathology from hospital biopsies. Non-dysplastic BE with no residual LGD was the diagnosis at visit 5.

The universally unmethylated profile of a number of target regions, such as Tubulin, alpha 3f, pseudogene (TUBA3FP) and ADP-ribosylation factor-like 10 (ARL10) (Figure 6-11) for a patient with persistent LGD (untreated, constant surveillance) is consistent with the profile of non-dysplastic BE patients. This methylation pattern was evident in a number of other target regions including calcium channel, voltage-dependent, alpha 2 / delta subunit 2 (CACNA2D2), leucine rich repeat containing 43 (LRRC43) and tetratricopeptide repeat and ankyrin repeat containing 1 (TRANK1) (data not shown). On follow-up, this patient was diagnosed as being without dysplasia two years after our LGD biopsy was taken and analyzed.

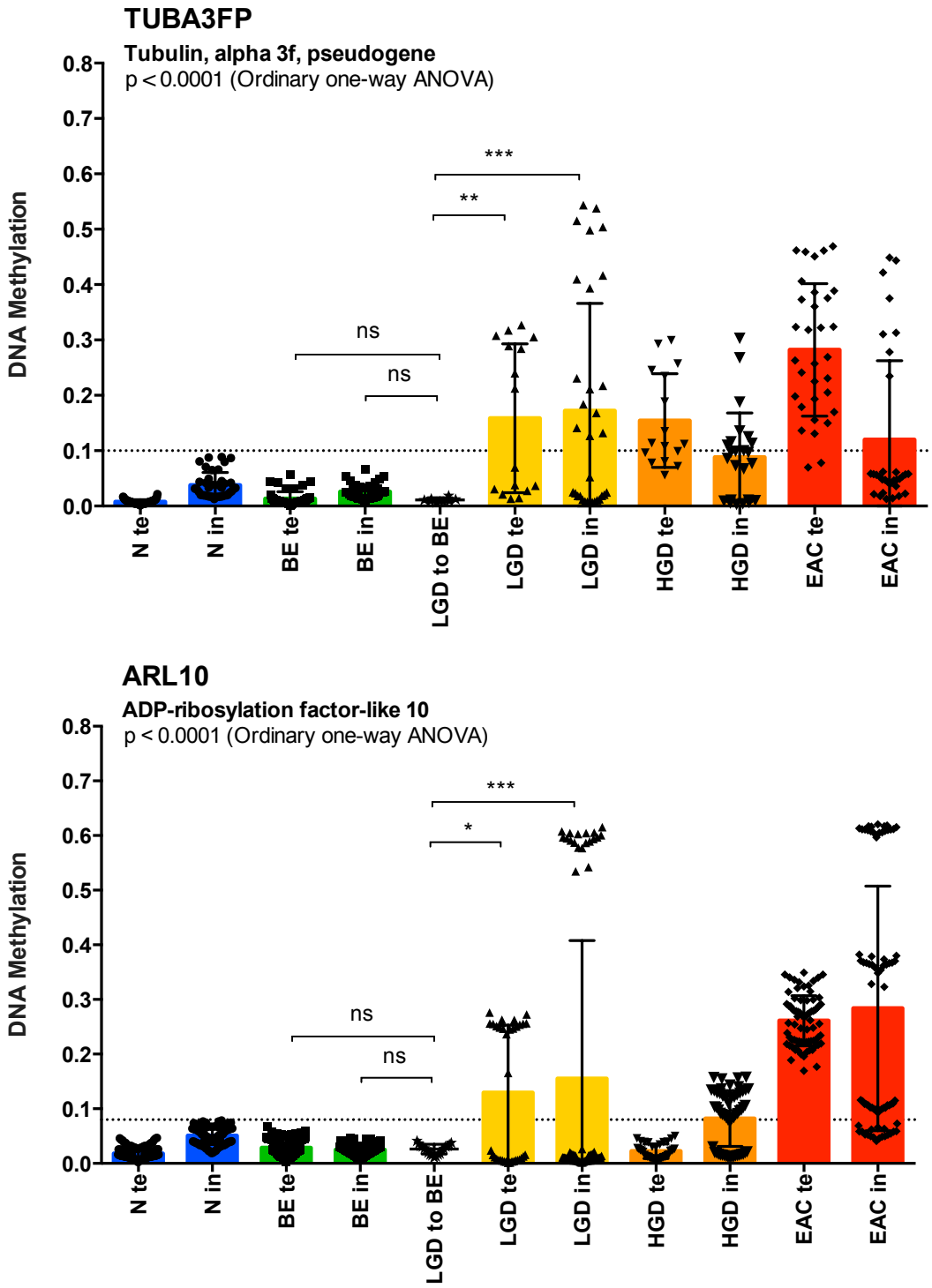


Figure 6-11: Methylation biomarker target regions TUBA3FP and ARL10 for predicting the necessity of treatment for low-grade dysplasia. Average methylation at each CpG site within the target region (TUBA3FP 8 sites, ARL10 19 sites) plotted for all patients in both the technical (te) and independent (in) validation cohorts (total of 48 samples from 33 patients). Statistical significance was evaluated using an ordinary one-way ANOVA test

and Dunnett's multiple comparisons test, with a single pooled variance. ns = not significant.

These results provide evidence that absence of methylation in these target regions in LGD patients may indicate predisposition to natural disease regression, thus warranting a less aggressive treatment or surveillance approach. A full study over an elongated time period with increased numbers of patients with is needed to confirm this hypothesis. This patient has been flagged for follow up: DNA has been extracted from a tissue sample taken at visit 5 (non-dysplastic BE diagnosis) and sent to our collaborators in Queensland (Dr Andrew Barbour) for HM450 profiling. This data has been included on a collaborative NHMRC grant application to investigate the progression of BE to dysplastic disease and adenocarcinoma.

6.3.6 Methylation biomarkers for disease monitoring with treatment

There is evidence of application of these biomarkers for monitoring disease regression with treatment. Included in the independent validation cohort were multiple samples taken from a single patient who was successfully treated. The patient was monitored over three visits in the space of 7½ months. A baseline methylation profile of matched normal esophageal tissue was taken at the first visit (proximally at 25cm), and diseased tissue samples at this (HGD) and each subsequent (LGD, then non-dysplastic BE) visit were taken prior to treatment.

The methylation profile of tumor necrosis factor, alpha-induced protein 8-like 3 (TNFAIP8L3) and zinc finger protein 699 (ZNF699) showed significant correlation ($p=0.0019$, $p<0.0001$ respectively) with disease regression over time as the patient was treated. Results are the average methylation detected at each CpG in the target region using MiSeq amplicon sequencing (Figure 6-12).

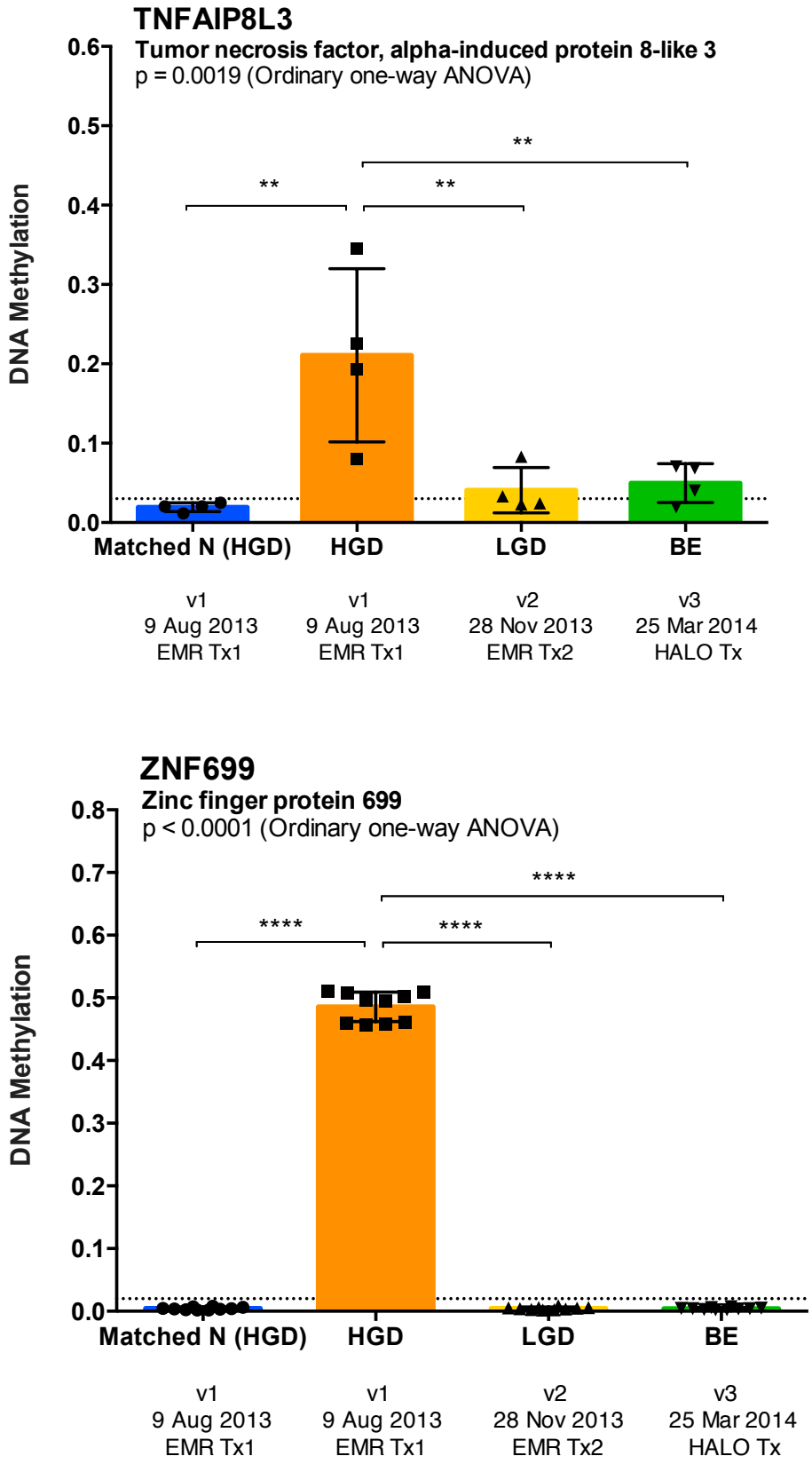


Figure 6-12: Methylation biomarker target regions TNFAIP8L3 and ZNF699 for monitoring disease regression with treatment. Statistical significance was

evaluated using an ordinary one-way ANVOA test and Dunnett's multiple comparisons test, with single pooled variance. Tx: treatment, v: visit.

Interestingly both markers showed a significant drop in methylation by the second visit, despite the patient still being classified as having LGD. Following this analysis the patient has had two further visits and was diagnosed with non-dysplastic BE on both occasions. Methylation of these target regions may be clinically valuable for monitoring disease status and with treatment and possibly have application in predicting disease recurrence as part of a screening panel.

6.4 Discussion

The results presented in this chapter give insight into methylation biomarker performance (both individually and as a panel) and possible utility of the identified biomarkers for diagnosis, prediction of progression and disease monitoring. Irrespective of comparison group, cut-off criteria for biomarker selection enforced a minimal methylation, or unmethylated status in normal healthy esophageal tissue, duodenal and proximal stomach tissue and normal peripheral blood from healthy individuals, with biomarkers showing disease-associated hypermethylation within specified target regions. The mucosal structure of proximal stomach and duodenal tissue controls, similar to that of Barrett's esophagus, enabled filtering of hypermethylation attributable to the maintenance of columnar epithelium, leaving only disease-associated aberrant methylation.

The N v BE comparison class identified biomarkers hypermethylated in non-dysplastic Barrett's epithelium with respect to normal, healthy esophageal mucosa. Typically, CpG sites identified by this comparison were very highly hypermethylated in Barrett's epithelium (up to 80%, with average baseline methylation of normal epithelium typically 1-2%). The unmethylated status of duodenal and proximal stomach control tissues means that these biomarkers specifically indicate the presence of Barrett's esophagus. No restrictions were placed on methylation in dysplastic BE or EAC for this comparison,

however interestingly, hypermethylation biomarkers for non-dysplastic BE remained highly methylated in dysplastic BE and EAC, indicating that disease-associated hypermethylation occurring early is maintained throughout metaplasia-dysplasia-adenocarcinoma progression. These biomarkers may be useful for confirming the presence of BE, however do not offer any clinical advantage over current histological identification, thus the focus of this thesis remained on the clinically relevant BE v HGD-EAC comparison.

The N v HGD-EAC comparison class identified biomarkers hypermethylated in intervention requiring disease in comparison to normal, healthy esophageal epithelium. In general, these CpG sites were very highly methylated in advanced disease, with average hypermethylation of up to 80% and average baseline methylation typically ~1-2%. No restrictions were placed on methylation in non-dysplastic BE and as such, methylation varied, but generally showed an intermediate level of hypermethylation at the sites that were highly hypermethylated in advanced disease. The clinical utility of these biomarkers is debatable. Detection and ongoing surveillance of non-dysplastic BE, with its high prevalence (estimated incidence of 1-2% of the adult population[34]) and very low progression rate (estimated <0.5% per year[39]), is not economically viable for the healthcare system. Until current endoscopic / histologic surveillance programs for BE evolve, the focus for early detection needs to remain on identification of dysplastic BE and EAC. The observation that >70% of hypermethylation biomarkers for the identification of BE overlap with those for HGD-EAC detection supports observations by Krause et al[49], that aberrant methylation occurring early in BE transformation is maintained throughout EAC development.

The BE v HGD-EAC comparison class identified biomarkers hypermethylated in intervention requiring disease with respect to non-dysplastic Barrett's and normal, healthy esophageal epithelium. In general, these sites were not as highly methylated as those identified in comparisons with normal esophageal epithelium only, however did exhibit hypermethylation up to 60% with

average baseline methylation typically ~1-2%. This is the most clinically valuable comparison class, distinguishing patients that require intervention from those that do not. Despite exclusion of the difficult to classify low-grade dysplastic tissues from algorithms to identify these biomarkers, a number of the selected target regions also showed good sensitivity (up to 85%) for disease detection in these samples.

With regards to differentiation of intervention requiring disease from non-dysplastic Barrett's epithelium, sensitivity and specificity of validated target regions was generally insufficient for consideration as stand-alone biomarkers. However, multiplexing target regions into panels for identification of early-stage disease results in significant improvement in performance. A triplex of three target regions (TUBA3FP, VANGL2 and ARL10) is able to successfully differentiate high-grade dysplasia and EAC from normal and BE samples with 100% sensitivity and 84.6% specificity (4 false positives in 26 baseline samples). For the identification of any dysplastic disease, including the notoriously difficult to diagnose LGD patients (differentiating LGD, HGD and EAC from normal and BE patients), a triplex of TUBA3FP, ARL10 and ZNF699 reported 95% sensitivity (a single LGD patient unable to be identified) and 76.9% specificity (6 false positives in 26 baseline samples).

By performing comprehensive clinical follow-up on all patients that had associated targeted sequencing data, I have demonstrated the possible utility of these biomarkers for the prediction of progression from non-dysplastic to dysplastic disease. Particularly remarkable was that increased methylation of target regions TEPP (testis, prostate and placenta expressed) and upstream of CA4 (carbonic anhydrase IV) was observed in a non-dysplastic BE sample, one-year prior to dysplastic progression.

Results were also suggestive of biomarker utility for prediction of necessity of treatment for LGD, a heavily debated topic in guidelines worldwide[52, 81]. A patient with persistent LGD (untreated, classification confirmed over several visits) showed minimal methylation (<0.05; as for non-dysplastic and normal

healthy esophagus) in several target regions (including TUBA3FP, ARL10, CACNA2D2, LRRC43 and TRANK1), 2 years prior to natural regression (without treatment) of disease to non-dysplastic status. Conversely, there is also evidence of application of these biomarkers for prediction of low to high-grade dysplasia. A patient with persistent long-segment LGD showed elevated methylation in LRRC43 and ZNF699, eight months prior to progression to HGD, despite several endoscopic mucosal resection treatments. In particular, hypermethylation of the regulatory region associated with leucine rich repeat containing 43 (LRRC43) shows evidence of utility for prediction of necessity of treatment for LGD; predicting both a natural regression of disease without treatment, as well as disease progression despite treatment.

Finally, several biomarkers (in particular, TNFAIP8L3 and ZNF699) also showed evidence of application for monitoring disease regression with treatment. Elevated methylation levels apparent in HGD samples show return to unmethylated status (as for matched normal esophagus) in non-dysplastic BE, with successful ablation therapy.

In summary, I have demonstrated the potential of selected hypermethylation biomarkers for diagnostic application, prediction of disease progression or regression (and hence an aid for determining treatment and/or surveillance regimes for LGD patients), and application for disease monitoring with treatment.

CHAPTER 7: PAN-CANCER MUTATION PROFILING

7.1 Introduction

7.1.1 Pan-cancer mutation panel development

Researchers at The Kinghorn Cancer Centre (TKCC), Sydney, Australia, are developing a novel pan-cancer mutation screening panel (not yet publicly available), aiming to provide a wider, all-encompassing analysis across disparate cancer types. As currently available panels tend to be specific to certain major cancers, this panel focuses on inclusion of genes to identify rare or poorly studied cancer types as well as screening for mutations in well-known cancer-associated genes. I would like to acknowledge Professor Marcel Dinger and his team at The Kinghorn Cancer Centre for including me in the development of this panel. For the collaboration, I provided his team with esophageal tissue DNA to aid in panel development; in return, they provided me with sequencing results for interrogation of the training cohort with regards to cancer-associated gene mutations. I would also like to acknowledge Dr Mark McCabe for his help in library preparation and processing of raw data.

7.1.2 Mutational landscape of esophageal adenocarcinoma

The molecular basis of events involved in BE progression to EAC are still debated with a large number of recurrently mutated genes in EAC also evident in non-dysplastic, non-progressive BE[206]. High mutational frequency is observed in EAC, but as yet no direct causality has been confirmed[207].

The genomic landscape of somatic alteration in esophageal adenocarcinoma has been the subject of much investigation over the past years. In an exome and whole-genome sequencing study in 2013, Dulak et al identified 26 significantly mutated genes from analysis of 149 EAC tumor-normal pairs, of which 21 were novel (TP53, CDKN2A, SMAD4, ARID1A and PIK3CA have been previously implicated in EAC)[208]. Investigation of somatic mutation, not only in EAC tissues, but also dysplastic and non-dysplastic Barrett's (EAC

precursor) mucosa was investigated by Weaver et al in 2014. Their approach involved identification of recurrently mutated genes in EAC followed by screening benign, metaplastic, never-dysplastic BE, and high-grade dysplasia. Interestingly they found that the majority of recurrently mutated genes in EAC were also mutated in non-dysplastic BE, with only TP53 and SMAD4 mutations confined to HGD and EAC (Figure 7-1)[206]. This is an interesting observation as it supports the hypothesis that EAC is not a mutation driven cancer, given that the majority (99.5%) of non-dysplastic BE cases will not progress to dysplasia or EAC.

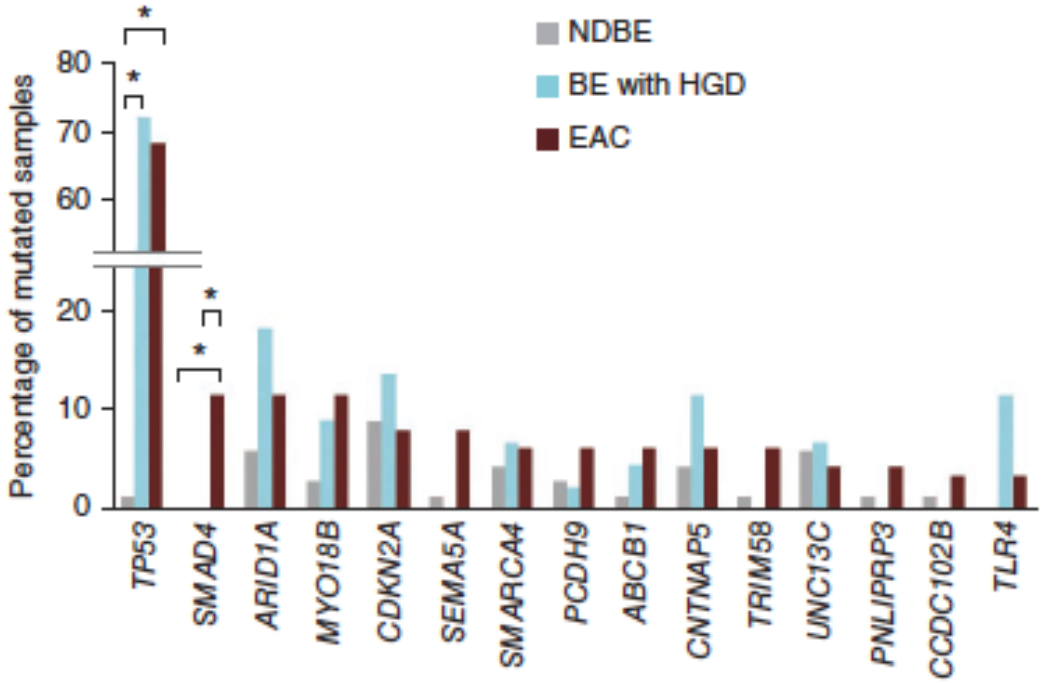


Figure 7-1: Percentage of never-dysplastic Barrett's esophagus (NDBE), high grade dysplasia (HGD) and esophageal adenocarcinoma (EAC) with mutations in recurrently mutated genes (mutated in ≥4 samples) occurring in the EAC discovery and validation cohorts. TP53 and SMAD4 are the only genes with mutations specific to HGD and EAC only ($p<0.05$). Figure from Weaver et al 2014[206].

Despite mounting evidence that many cases of EAC may not be mutation-driven, there are still a number of genes identified in literature as significantly

mutated in EAC that are analyzed as part of the Kinghorn pan-cancer mutation screening panel (Figure 7-2). Of particular interest are TP53 and SMAD4, specific to only intervention-requiring disease and not occurring in non-dysplastic precursor disease mucosa. The eight significantly mutated genes in EAC (ARID1A, ARID2, CDKN2A (P16), KAT6A, PIK3CA, SMAD4, SMARCA4, TP53) identified in the literature[206, 208] and also in the pan-cancer panel will be analyzed in training cohort samples (genome-wide methylation and expression profiling data exists for these same samples). Further mutations may also be identified.

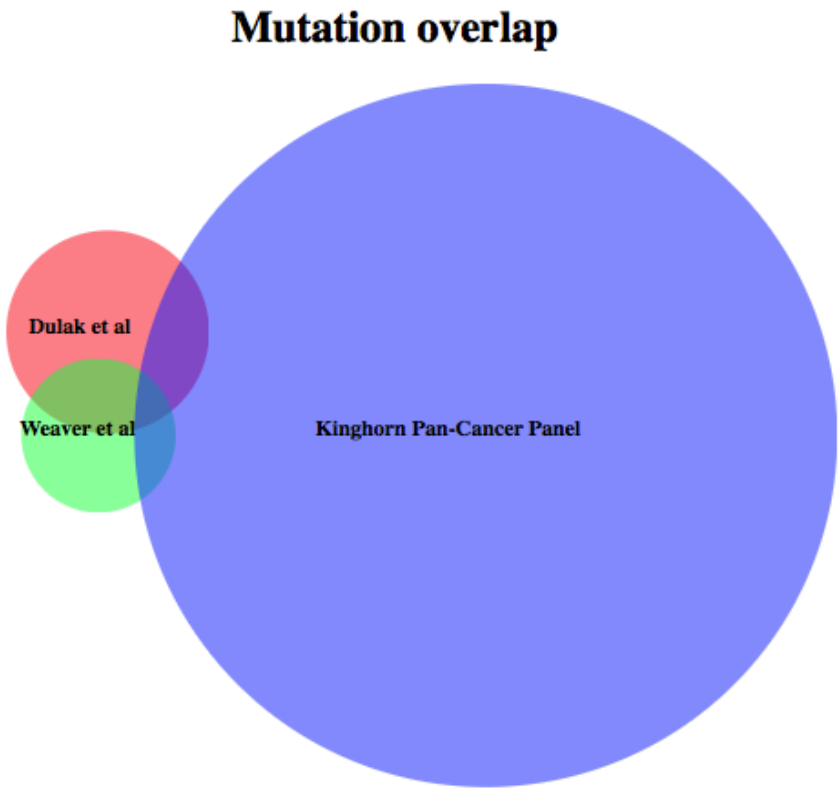


Figure 7-2: Overlap of significantly mutated genes in EAC identified by Dulak et al (n=26) and Weaver et al (n=15) with the Kinghorn pan-cancer panel for rare and poorly understood cancers (n=312). Seven genes were common between the two studies (ARID1A, CDKN2A, CNTNAP5, SMAD4, SMARCA4, TLR4, TP53), with eight of the Dulak genes in the Kinghorn panel (ARID1A, ARID2, CDKN2A, KAT6A, PIK3CA, SMAD4, SMARCA4, TP53) and five from the Weaver study in the Kinghorn panel (ARID1A,

CDKN2A, SMAD4, SMARCA4, TP53 (common to all)). Venn diagram created using BioVenn[205]

7.1.3 Somatic mutation calling

Somatic mutation calling from matched tumor-normal patient samples has become a critical part of cancer genomics, not only for characterization, but also clinical applications such as determining treatment regimens[209]. Tumor heterogeneity, copy number alteration and sample degradation present a challenge for the detection of somatic mutation, combined with base-calling error and read alignment problems to be overcome for robust mutation calling accuracy[210, 211]. Strelka, used here for somatic mutation calling, uses a joint probability-based statistical approach for simultaneous matched tumor-normal dataset analysis. It has been shown to achieve significantly higher sensitivity at the lowest SNV fraction, maintaining low false positive rates[209].

7.1.4 Molecular basis of CIMP in human neoplasia

CpG island methylator phenotype (CIMP), or widespread CpG island promoter hypermethylation, was initially identified in colorectal cancer, but has since been studied in a number of tumor types, including but not limited to breast, bladder, gastric, glioblastoma (gliomas), hepatocellular, lung, ovarian, pancreatic, prostate, melanoma and duodenal adenocarcinoma[212, 213].

Interestingly, the molecular basis of CIMP in glioma was shown to be established by, and highly dependent on, the mutation of a *single* gene: isocitrate dehydrogenase 1 (IDH1)[214]. The mutation, most commonly (>95% in glioma[214]) an amino acid substitution at arginine 132 (R132), alters specific histone marks, induces extensive DNA hypermethylation. Similarly in acute myeloid leukemia (AML), a casual relationship was identified between IDH1 and IDH2 mutations and global DNA hypermethylation (a specific hypermethylation signature). Furthermore, Tet methylcytosine dioxygenase 2 (TET2) loss-of-function mutations were mutually exclusive with IDH1/2

mutations, and also associated with similar epigenetic changes[215]. The prevalence of somatic IDH mutation among cancer types is highly variable, with rates varying from >70% in glioma, 15-30% in AML, 10% in melanoma and <5% or absent in many other solid tumors, such as gastrointestinal stromal tumors (GIST), bladder, breast, colorectal, lung, ovarian, pancreas, prostate and thyroid carcinoma[212, 216-218], despite CIMP being reported in these tumors[219-226].

The molecular basis of CIMP in colorectal cancer has been suggested to be one of two types: either v-raf murine sarcoma viral oncogene homolog B1 (BRAF) mutation (known as 'CIMP high') or v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS) mutation ('CIMP low')[227, 228]. This is somewhat surprising considering both KRAS and BRAF are members of the ras signal transduction pathway and are typically mutually exclusive[229]. The question of whether BRAF mutation is causal for CIMP development or whether CIMP provides a favorable environment for BRAF mutation acquisition was investigated by Hinoue et al in colorectal cancer cell lines. They did not find any evidence of a causal relationship upon examination of 100 CIMP-associated CpG sites in eight stably transfected clones over multiple passages[230].

It is evident that CIMP is not a single phenotype consistent across all cancer types. In a review in 2013, Hughes et al suggested a cancer-specific CIMP nomenclature be adopted, for example G-CIMP for glioma CIMP, etc...[212] Esophageal adenocarcinoma falls under the umbrella of gastrointestinal (GI: esophagus, stomach, small and large intestine, liver, gallbladder and pancreas) cancers, for which many epigenetic similarities exist[231]. For example, colorectal CIMP has been reported to be prevalent in a subset of hepatocellular carcinoma (HCC)[232]. However, despite their gastrointestinal grouping and mucosal structure similarities, there are also many epigenetic disparities between these cancers. There is evidence of E-CIMP (esophageal adenocarcinoma CIMP) being similar to CIMP as defined in gastric and colorectal cancers[49], however the molecular basis of CIMP in EAC is yet to be elucidated.

7.1.5 Chapter 7 aims

To ascertain cancer-associated mutational load in the development of esophageal adenocarcinoma, and use this information to gain a more complete understanding of genetic and epigenetic changes occurring in the metaplasia-dysplasia-adenocarcinoma sequence.

7.2 Methods

7.2.1 Pan-cancer mutation panel target selection

A novel custom tumor gene panel designed through Roche/NimbleGen by the Hormones and Cancer group at the Kinghorn Centre for Clinical Genomics at the Garvan Institute of Medical Research was applied to a subset of the training cohort. Dr Mark McCabe kindly performed the library preparation as well as processing of the raw sequencing data.

The panel uses capture sequencing to probe a set of >300 pan-cancer genes in their entirety (refer to Appendix 8: Pan-cancer targets for mutational profiling for a full list of genes examined by the panel); the selected targets based on general cancer-associated genes from multiple commercially available panels (Illumina's TruSeq Tumor Panel and TruSeq Amplicon Cancer Panel, Oxford Gene Technology's SureSeq Solid Tumor Panel, Foundation Medicine's FoundationOne panel, Agena Bioscience's OncoCarta volumes 1, 2 and 3 and Aglient's Haloplex panel). The panel is still under development and aims to address carcinogenic changes in rare and poorly understood cancers; using blood and tissue derived tumor DNA from pituitary, head and neck (salivary gland and skull base), breast (phyllodes), pancreas, lung and esophageal adenocarcinoma for development.

7.2.2 Library preparation and sequencing

100ng of DNA extracted from esophageal tissue (biopsies as for HM450 and HTA2.0 profiling) was subject to library preparation using a combination of the NEBNext Ultra kit (applicable to Illumina sequencing platforms) and

Roche's SeqCap Library SR Kit according to manufacturer's instructions. Capture sequencing of 24 samples was performed on a single lane on an Illumina HiSeq 2500.

7.2.3 Mutation-screening statistical analyses

Raw fastq files were uploaded to cloud-based genomic analysis platform DNAnexus (www.dnanexus.com). Sequences were genome-aligned to the hs37d5 reference genome using Burrows-Wheeler Aligner (BWA-MEM, v0.7.10), and then sorted with PCR duplicates marked using novosort (v1.03.01)[233]. Base quality (BQ) scores were generated for each base sequenced and used to predict the probability that the called base was the true base. Mapping quality (MQ) scores were assigned to each read; when MQ=0 (reads mapped to multiple locations within the genome), they were discarded. QUAL scores were assigned for each variant as a measure of its likelihood of being present in the cohort and finally a Genotype quality (GQ) score was obtained; measuring of the probability that the genotype called at the given allele in the patient was correct. QUAL and GQ scores incorporate BQ and MQ for each base and read that contains the given allele as well as the depth of coverage[234]. Aligned reads were processed according to the Genome Analysis Toolkit (GATK v3.3 (best practices guide)). Reads were realigned around indels and BQ scores recalibrated to improve the quality of alignments[235]. Single nucleotide variants and small insertions/deletions were identified using HaplotypeCaller v3.3 or Strekla (for somatic variant detection) and then annotated using Ensembl Variant Effect Predictor (VEP, v74)[236]. Annotation provides (i) chromosomal location of the variant, (ii) type of variant, (iii) predicted variant effect (benign or deleterious), (iv) occurrence in healthy individuals and (v) conservation of wild type alleles across multiple species. Pairwise analysis of each disease-normal pair was performed to identify disease-specific mutation. As for previous methylation and expression profiling studies, normal samples are esophageal tissue from patients with esophageal disease, with the sample taken proximally (at approximately 25cm). Analysis of all normal samples was performed to establish germline mutational load.

Data were filtered for high and medium impact variants. High impact variants include non-sense mutations (introduction of a stop codon), loss of a stop codon, splice acceptor or donor variant and any frame-shift variant. Most (if not all) missense mutation (non-synonymous change) will be classified as medium impact, regardless of subsequent amino acid change. Small in-frame insertions or deletions are classified as medium impact. The remainder are low impact, including intronic variants, UTR variants and silent mutations (synonymous variants; DNA change but no impact on protein sequence).

7.3 Results

7.3.1 Pan-cancer panel profiling

Sequencing was of consistently high coverage, with >90% of targeted regions at >100x, with average depth of coverage ~700x. Analysis of over 4372 exons in 313 cancer-associated genes enabled detection of germline and somatic mutations in esophageal tissue as disease develops and progresses.

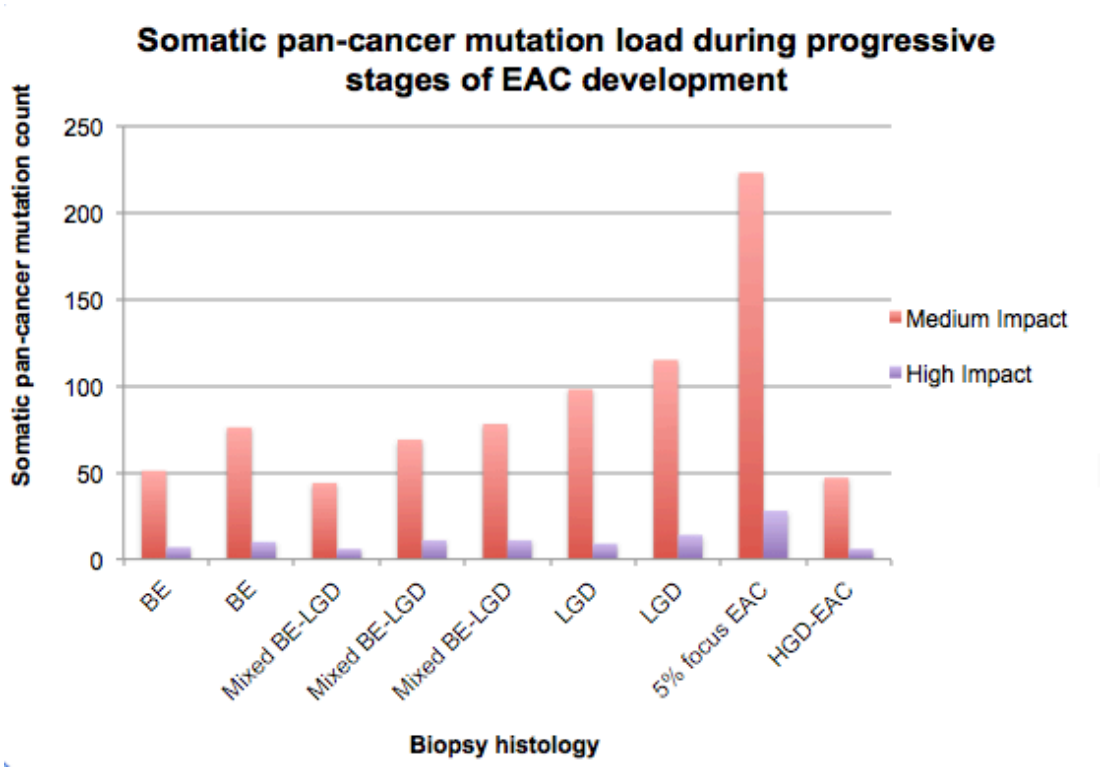
7.3.1.1 Mutational load

Normal samples (n=15) were analyzed to establish germline mutational load. Somatic pan-cancer mutation load was calculated for each of 9 disease-normal pairs throughout the metaplasia-dysplasia-adenocarcinoma sequence (BE (n=2), BE-LGD mixed (n=3), LGD (n=2), EAC-HGD (n=2)). Results were separated into mutations classified as either high or medium impact (by Variant Effect Predictor (VEP)). Mutational load varied, but in general an increase in mutational load was detected as disease progressed (Figure 7-3). It is interesting to note the detection of cancer-associated mutation in Barrett's mucosa, a non-dysplastic, benign epithelium. The detection of mutations with a cancer association at this early pre-cancerous stage indicates that genetic aberration is occurring early in EAC carcinogenesis.

An interesting anomaly is the distinct lack of cancer associated somatic mutation in one of the EAC samples (labeled HGD-EAC in Figure 7-3 below). In comparison to a biopsy containing just a 5% focus of destroyed glands

suspected to be adenocarcinoma (refer to Section 4.3.3 Tissue heterogeneity), mutational load for this sample is significantly reduced (4.74 fold decrease in high and medium impact mutations). The sample was taken from a 70 year old male with invasive adenocarcinoma, one month before undergoing esophagectomy with the biopsy evaluated to be >80% EAC and noted as developed tumor by at least one pathologist. This anomaly highlights the complexities of EAC and the heterogeneous biological signatures within the disease. Detection of such high mutational load in a biopsy with only a 5% focus of adenocarcinoma reflects well on method sensitivity. We could also hypothesize that there may be some field effect where normal tissue surrounding neoplastic tissue becomes similarly mutated and thus is at risk of neoplastic transformation.

(A)



(B)

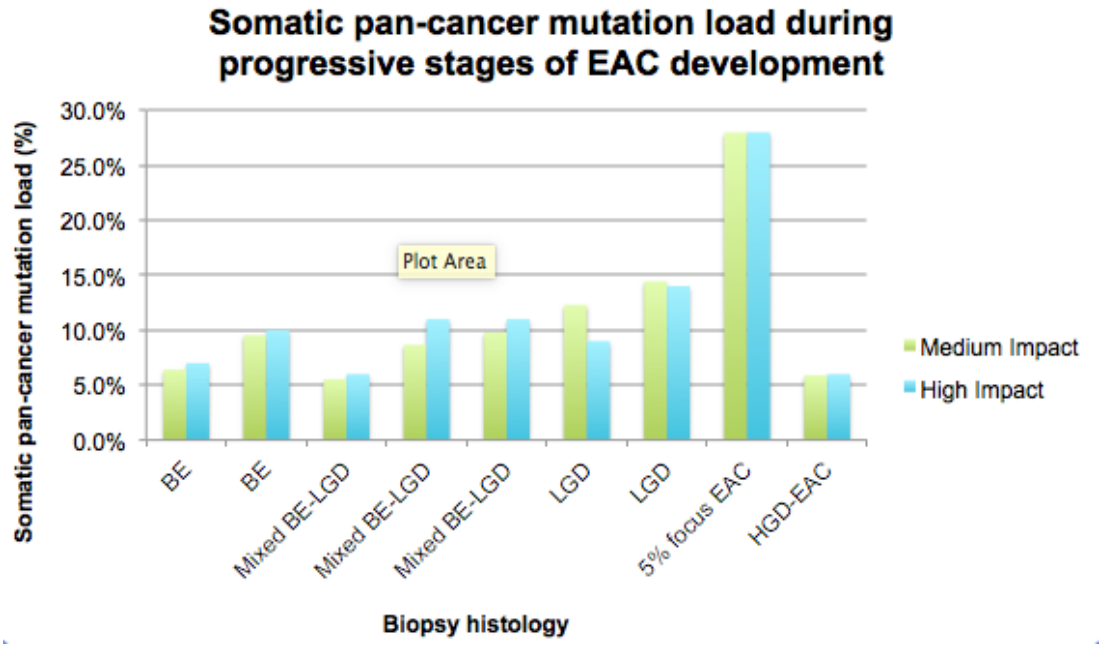


Figure 7-3: Somatic pan-cancer mutation load (high and medium impact mutations) detected in esophageal mucosa during progression through the metaplasia-dysplasia-adenocarcinoma sequence. (A) Mutation count in diseased samples (B) Expressed as a percentage of all mutations detected in esophageal tissue. Percentages are of 897 total SNPs/indels (100 high impact, 797 medium impact mutations). Somatic variants called by Strelka, variants annotated by Variant Effect Predictor (VEP), filtering performed using Gemini.

7.3.1.2 Disease-normal pairwise analyses

Pairwise-analysis of somatic mutation in 9 disease-normal pairs (BE (n=2), BE-LGD mixed (n=3), LGD (n=2), EAC-HGD (n=2)), classified as either high or medium impact (by VEP) uncovered 897 SNPs/indels (classifying 100 as high, 797 as medium impact mutations). Somatic variants called by Strelka, variants annotated by Variant Effect Predictor (VEP), filtering performed using Gemini.

Combined Annotation Dependent Depletion (CADD) is a tool for scoring the deleteriousness of SNVs and indels in the human genome, quantitatively

prioritizing functional, deleterious and disease causal variants.[237]. Applying minimum scaled CADD scoring of 20 (resulting in the top 1% most deleterious mutations) reduces this to identification of a total of 636 SNPs/indels (97 high and 539 medium impact mutations). Applying CADD 30 (resulting in the top 0.1% most deleterious mutations) further reduces this to identification of a total of 192 SNPs/indels (66 high and 126 medium impact mutations), given in Table 7-1.

Table 7-1: Top 0.1% most deleterious pan-cancer somatic mutation from pairwise disease-normal analysis of samples at all stages of esophageal adenocarcinoma development. Returning SNPs and indels classified as either high or medium impact by VEP. Minimum scaled CADD score 30. Ordered by impact severity then variant location.

Variant	Type	Gene	Impact	Impact Severity
chr1:g.16255388G>T	SNP	SPEN	stop_gained	HIGH
chr1:g.17349186C>A	SNP	SDHB	stop_gained	HIGH
chr1:g.27057682C>T	SNP	ARID1A	stop_gained	HIGH
chr1:g.27106373AGGCAATGACTT TGAGATGTCCAAACACCCAG>A	Deletion	ARID1A	frameshift_variant	HIGH
chr1:g.27106861C>T	SNP	ARID1A	stop_gained	HIGH
chr1:g.206649604G>T	SNP	IKBKE	stop_gained	HIGH
chr11:g.113075226C>T	SNP	NCAM1	stop_gained	HIGH
chr12:g.46287298C>A	SNP	ARID2	stop_gained	HIGH
chr16:g.3779388G>C	SNP	CREBBP	stop_gained	HIGH
chr16:g.3789709C>A	SNP	CREBBP	stop_gained	HIGH
chr16:g.3828718C>A	SNP	CREBBP	stop_gained	HIGH
chr16:g.23641437C>A	SNP	PALB2	stop_gained	HIGH
chr16:g.68772290G>T	SNP	CDH1	stop_gained	HIGH
chr16:g.89880936G>C	SNP	FANCA	stop_gained	HIGH
chr17:g.7578263G>A	SNP	TP53	stop_gained	HIGH
chr17:g.37868607G>T	SNP	ERBB2	stop_gained	HIGH
chr17:g.41245042C>A	SNP	BRCA1	stop_gained	HIGH
chr17:g.49244258G>T	SNP	NME1-NME2	stop_gained	HIGH
chr17:g.59878695G>T	SNP	BRIP1	stop_gained	HIGH
chr19:g.10597459C>A	SNP	KEAP1	stop_gained	HIGH
chr19:g.10599868C>A	SNP	KEAP1	stop_gained	HIGH
chr19:g.11105636G>T	SNP	SMARCA4	stop_gained	HIGH
chr19:g.17954641T>TC	Insertion	JAK3	frameshift_variant	HIGH
chr2:g.141243069T>TGG	Insertion	LRP1B	frameshift_variant	HIGH
chr2:g.191905853C>A	SNP	STAT4	stop_gained	HIGH
chr2:g.234676519C>T	SNP	UGT1A10	stop_gained	HIGH
chr20:g.54956548C>A	SNP	AURKA	stop_gained	HIGH
chr22:g.41536182T>A	SNP	EP300	stop_gained	HIGH
chr22:g.41548270A>T	SNP	EP300	stop_gained	HIGH
chr3:g.36779163C>A	SNP	DCLK3	stop_gained	HIGH
chr3:g.37042446A>T	SNP	MLH1	stop_gained	HIGH
chr3:g.52643906C>A	SNP	PBRM1	stop_gained	HIGH
chr3:g.138665322G>T	SNP	FOXL2	stop_gained	HIGH

Variant	Type	Gene	Impact	Impact Severity
chr3:g.142280158C>A	SNP	ATR	stop_gained	HIGH
chr4:g.1806083G>T	SNP	FGFR3	stop_gained	HIGH
chr4:g.66286164C>A	SNP	EPHA5	stop_gained	HIGH
chr4:g.104072462C>A	SNP	CENPE	stop_gained	HIGH
chr4:g.104081828G>T	SNP	CENPE	stop_gained	HIGH
chr4:g.106156095C>A	SNP	TET2	stop_gained	HIGH
chr4:g.153271230A>T	SNP	FBXW7	stop_gained	HIGH
chr5:g.41929932G>T	SNP	FBXO4	stop_gained	HIGH
chr5:g.56178614C>G	SNP	MAP3K1	stop_gained	HIGH
chr5:g.64766668G>T	SNP	ADAMTS6	stop_gained	HIGH
chr5:g.98236518C>A	SNP	CHD1	stop_gained	HIGH
chr5:g.112128218G>T	SNP	APC	stop_gained	HIGH
chr5:g.112175525G>T	SNP	APC	stop_gained	HIGH
chr5:g.131939047C>T	SNP	RAD50	stop_gained	HIGH
chr5:g.149440518C>A	SNP	CSF1R	stop_gained	HIGH
chr5:g.180058728C>A	SNP	FLT4	stop_gained	HIGH
chr6:g.393348G>T	SNP	IRF4	stop_gained	HIGH
chr7:g.55229238C>A	SNP	EGFR	stop_gained	HIGH
chr7:g.81335691C>A	SNP	HGF	stop_gained	HIGH
chr7:g.140453092C>A	SNP	BRAF	stop_gained	HIGH
chr7:g.151835883C>A	SNP	KMT2C	stop_gained	HIGH
chr7:g.151845205C>A	SNP	KMT2C	stop_gained	HIGH
chr7:g.151874215C>A	SNP	KMT2C	stop_gained	HIGH
chr8:g.37698908C>T	SNP	GPR124	stop_gained	HIGH
chr9:g.37015184C>A	SNP	PAX5	stop_gained	HIGH
chr9:g.98268881C>A	SNP	PTCH1	initiator_codon_variant	HIGH
chr9:g.135771912C>A	SNP	TSC1	stop_gained	HIGH
chr9:g.139393640GC>G	Deletion	NOTCH1	frameshift_variant	HIGH
chr9:g.139413062C>CT	Insertion	NOTCH1	frameshift_variant	HIGH
chrX:g.53253957C>A	SNP	KDM5C	stop_gained	HIGH
chrX:g.63410150G>T	SNP	AMER1	stop_gained	HIGH
chrX:g.76889095C>A	SNP	ATRX	stop_gained	HIGH
chrX:g.123200089G>T	SNP	STAG2	stop_gained	HIGH
chr1:g.11167547A>T	SNP	MTOR	missense_variant	MED
chr1:g.11187790C>A	SNP	MTOR	missense_variant	MED
chr1:g.38005813C>A	SNP	SNIP1	missense_variant	MED
chr1:g.65313346C>A	SNP	JAK1	missense_variant	MED
chr1:g.120464359C>A	SNP	NOTCH2	missense_variant	MED
chr1:g.150549919T>A	SNP	MCL1	missense_variant	MED
chr1:g.186645981C>A	SNP	PTGS2	missense_variant	MED
chr1:g.228111919C>A	SNP	WNT9A	missense_variant	MED
chr1:g.243716220C>A	SNP	AKT3	missense_variant	MED
chr1:g.243828074C>A	SNP	AKT3	missense_variant	MED
chr10:g.123263433G>T	SNP	FGFR2	missense_variant	MED
chr11:g.31823123C>A	SNP	PAX6	missense_variant	MED
chr11:g.32439161C>T	SNP	WT1	missense_variant	MED
chr11:g.76158044G>C	SNP	C11orf30	missense_variant	MED
chr11:g.76175054C>A	SNP	C11orf30	missense_variant	MED
chr11:g.94180495G>T	SNP	MRE11A	missense_variant	MED
chr11:g.94212841C>A	SNP	MRE11A	missense_variant	MED
chr11:g.108190782C>A	SNP	ATM	missense_variant	MED
chr11:g.118342932G>T	SNP	KMT2A	missense_variant	MED
chr11:g.118360543C>A	SNP	KMT2A	missense_variant	MED
chr11:g.118361991C>A	SNP	KMT2A	missense_variant	MED
chr11:g.118392014A>T	SNP	KMT2A	missense_variant	MED
chr11:g.125496668C>A	SNP	CHEK1	missense_variant	MED
chr12:g.441056G>A	SNP	KDM5A	missense_variant	MED
chr14:g.38061763C>A	SNP	FOXA1	missense_variant	MED

Variant	Type	Gene	Impact	Impact Severity
chr14:g.62194244C>A	SNP	HIF1A	missense_variant	MED
chr14:g.62199176G>T	SNP	HIF1A	missense_variant	MED
chr14:g.62204846T>A	SNP	HIF1A	missense_variant	MED
chr14:g.105239678C>A	SNP	AKT1	missense_variant	MED
chr15:g.40993283G>A	SNP	RAD51	missense_variant	MED
chr15:g.88680635C>A	SNP	NTRK3	missense_variant	MED
chr15:g.90630708C>A	SNP	IDH2	missense_variant	MED
chr16:g.3827632G>A	SNP	CREBBP	missense_variant	MED
chr16:g.9916227C>G	SNP	GRIN2A	missense_variant	MED
chr16:g.9934874C>A	SNP	GRIN2A	missense_variant	MED
chr16:g.23625399C>G	SNP	PALB2	missense_variant	MED
chr16:g.67645870C>A	SNP	CTCF	missense_variant	MED
chr16:g.67660599G>T	SNP	CTCF	missense_variant	MED
chr16:g.67663314G>T	SNP	CTCF	missense_variant	MED
chr17:g.7577105G>A	SNP	TP53	missense_variant	MED
chr17:g.7577121G>A	SNP	TP53	missense_variant	MED
chr17:g.7578526C>A	SNP	TP53	missense_variant	MED
chr17:g.8108544C>A	SNP	AURKB	missense_variant	MED
chr17:g.29683562C>A	SNP	NF1	missense_variant	MED
chr17:g.37618417AGCAGAG>A	Deletion	CDK12	inframe_deletion	MED
chr17:g.37872610G>T	SNP	ERBB2	missense_variant	MED
chr17:g.37880201C>A	SNP	ERBB2	missense_variant	MED
chr17:g.38548597TA>T	Deletion	TOP2A	splice_region_variant	MED
chr17:g.49244270C>T	SNP	NME1-NME2	missense_variant	MED
chr17:g.59857626G>A	SNP	BRIP1	missense_variant	MED
chr17:g.63010814C>A	SNP	GNA13	missense_variant	MED
chr17:g.63052690G>A	SNP	GNA13	missense_variant	MED
chr18:g.45374860G>T	SNP	SMAD2	missense_variant	MED
chr18:g.48591925G>T	SNP	SMAD4	missense_variant	MED
chr19:g.11101922C>T	SNP	SMARCA4	missense_variant	MED
chr19:g.11170507G>T	SNP	SMARCA4	missense_variant	MED
chr19:g.17942116C>A	SNP	JAK3	missense_variant	MED
chr19:g.17942163G>T	SNP	JAK3	missense_variant	MED
chr19:g.30311640C>A	SNP	CCNE1	missense_variant	MED
chr2:g.29445210C>A	SNP	ALK	missense_variant	MED
chr2:g.29456471C>A	SNP	ALK	missense_variant	MED
chr2:g.39240649C>A	SNP	SOS1	missense_variant	MED
chr2:g.39283916C>A	SNP	SOS1	missense_variant	MED
chr2:g.58387256G>T	SNP	FANCL	missense_variant	MED
chr2:g.61717804C>A	SNP	XPO1	missense_variant	MED
chr2:g.121729577C>T	SNP	GLI2	missense_variant	MED
chr2:g.140990850C>A	SNP	LRP1B	missense_variant	MED
chr2:g.141291595C>A	SNP	LRP1B	missense_variant	MED
chr2:g.141751626C>A	SNP	LRP1B	missense_variant	MED
chr2:g.170059434C>A	SNP	LRP2	missense_variant	MED
chr2:g.170092360G>T	SNP	LRP2	missense_variant	MED
chr2:g.170099933C>A	SNP	LRP2	missense_variant	MED
chr2:g.170150660C>A	SNP	LRP2	missense_variant	MED
chr2:g.191940984C>A	SNP	STAT4	missense_variant	MED
chr2:g.198257857C>A	SNP	SF3B1	missense_variant	MED
chr2:g.212543853C>A	SNP	ERBB4	missense_variant	MED
chr2:g.212812239C>A	SNP	ERBB4	missense_variant	MED
chr20:g.36026228G>T	SNP	SRC	missense_variant	MED
chr20:g.36031164C>A	SNP	SRC	missense_variant	MED
chr20:g.36031585C>T	SNP	SRC	missense_variant	MED
chr20:g.39744957G>T	SNP	TOP1	missense_variant	MED
chr20:g.62331856C>A	SNP	ARFRP1	missense_variant	MED
chr22:g.30035192C>A	SNP	NF2	missense_variant	MED

Variant	Type	Gene	Impact	Impact Severity
chr22:g.42522724G>T	SNP	CYP2D6	missense_variant	MED
chr3:g.10084295C>A	SNP	FANCD2	missense_variant	MED
chr3:g.12633289C>A	SNP	RAF1	missense_variant	MED
chr3:g.37089065G>T	SNP	MLH1	missense_variant	MED
chr3:g.47098879C>T	SNP	SETD2	missense_variant	MED
chr3:g.87313662C>A	SNP	POU1F1	missense_variant	MED
chr3:g.89468478C>A	SNP	EPHA3	missense_variant	MED
chr3:g.119631596C>T	SNP	GSK3B	missense_variant	MED
chr3:g.134911541C>T	SNP	EPHB1	missense_variant	MED
chr3:g.134920478C>A	SNP	EPHB1	missense_variant	MED
chr3:g.142184067A>G	SNP	ATR	missense_variant	MED
chr3:g.142218513C>A	SNP	ATR	missense_variant	MED
chr4:g.55962447C>G	SNP	KDR	missense_variant	MED
chr4:g.153245393C>A	SNP	FBXW7	missense_variant	MED
chr5:g.38966777C>A	SNP	RICTOR	missense_variant	MED
chr5:g.98233011C>A	SNP	CHD1	missense_variant	MED
chr5:g.138223183G>T	SNP	CTNNA1	missense_variant	MED
chr6:g.106543578C>A	SNP	PRDM1	missense_variant	MED
chr6:g.152382133G>A	SNP	ESR1	missense_variant	MED
chr6:g.152382257C>A	SNP	ESR1	missense_variant	MED
chr7:g.2977548C>A	SNP	CARD11	missense_variant	MED
chr7:g.2985588G>A	SNP	CARD11	missense_variant	MED
chr7:g.92354975G>A	SNP	CDK6	missense_variant	MED
chr7:g.116418839A>T	SNP	MET	missense_variant	MED
chr7:g.148516764G>T	SNP	EZH2	missense_variant	MED
chr7:g.151836816C>A	SNP	KMT2C	missense_variant	MED
chr7:g.151842295G>T	SNP	KMT2C	missense_variant	MED
chr7:g.151864233G>T	SNP	KMT2C	missense_variant	MED
chr8:g.37690545C>T	SNP	GPR124	missense_variant	MED
chr8:g.48811106C>G	SNP	PRKDC	missense_variant	MED
chr8:g.61777584C>A	SNP	CHD7	missense_variant	MED
chr9:g.5081832C>A	SNP	JAK2	missense_variant	MED
chr9:g.5089710C>A	SNP	JAK2	missense_variant	MED
chr9:g.87636192G>A	SNP	NTRK2	missense_variant	MED
chr9:g.98224263C>A	SNP	PTCH1	missense_variant	MED
chr9:g.135797319C>A	SNP	TSC1	missense_variant	MED
chrX:g.39916480C>A	SNP	BCOR	missense_variant	MED
chrX:g.44918322C>A	SNP	KDM6A	missense_variant	MED
chrX:g.47430356G>T	SNP	ARAF	missense_variant	MED
chrX:g.63412094C>A	SNP	AMER1	missense_variant	MED
chrX:g.76872147C>A	SNP	ATRX	missense_variant	MED
chrX:g.100609675C>A	SNP	BTK	missense_variant	MED
chrX:g.110385421C>A	SNP	PAK3	missense_variant	MED

OMIM® (Online Mendelian Inheritance in Man) was used for annotation of somatic mutation with human genetic disorders (http://omim.org/ Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, John Hopkins University, Baltimore, MD, USA). Table 7-2 gives the subset of top 0.1% most deleterious somatic mutations (applying CADD scoring of 30) with cancer-associated OMIM disorders (including cancer susceptibility). A total of 53 SNPs (20 high and 33 medium impact

mutations) meet this criteria. No mutations with esophageal adenocarcinoma OMIM disorder were detected in our samples, however it is important to note that esophageal mutations had not been specifically targeted during pan cancer panel design at this stage and our collaboration was intended to investigate utility of this panel across a variety of rare and poorly understood cancer types.

Table 7-2: Cancer-associated OMIM disorders from the subset of the top 0.1% most deleterious pan-cancer somatic mutation from pairwise disease-normal analysis of samples at all stages of esophageal adenocarcinoma development. Includes only somatic mutation with human cancer or cancer susceptibility OMIM annotation. Returning SNPs and indels classified as either high or medium impact by VEP. Minimum scaled CADD score 20. Mutations with gastrointestinal cancer (or susceptibility to) OMIM annotation in grey.

Gene	OMIM Title	OMIM Disorders
SPEN	SPEN, Drosophila, homolog of; One-twenty two protein	613484:None;606077:Megakaryoblastic leukemia, acute (2)
SDHB	Succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	185470:Paragangliomas 4, 115310 (3);Pheochromocytoma, 171300 (3);Paraganglioma and gastric stromal sarcoma, 606864 (3);Cowden syndrome 2, 612359 (3);Gastrointestinal stromal tumor, 606764 (3)
PALB2	Partner and localizer of BRCA2	610355:Fanconi anemia, complementation group N, 610832 (3);{Breast cancer, susceptibility to}, 114480 (3);{Pancreatic cancer, susceptibility to, 3}, 613348 (3)
CDH1	Cadherin-1 (E-cadherin; uvomorulin)	192090:Endometrial carcinoma, somatic, 608089 (3);Ovarian carcinoma, somatic, 167000 (3);{Breast cancer, lobular}, 114480 (3);Gastric cancer, familial diffuse, with or without cleft lip and/or palate, 137215 (3);{Prostate cancer, susceptibility to}, 176807 (3)
TP53	Tumor protein p53	191170:Colorectal cancer, 114500 (3);Li-Fraumeni syndrome, 151623 (3);Hepatocellular carcinoma, 114550 (3);Osteosarcoma, 259500 (3);Choroid plexus papilloma, 260500 (3);Nasopharyngeal carcinoma, 607107 (3);Pancreatic cancer, 260350 (3);Adrenal cortical carcinoma, 202300 (3);Breast cancer, 114480 (3);{Basal cell carcinoma 7}, 614740 (3);{Glioma susceptibility 1}, 137800 (3)
ERBB2	Avian erythroblastic leukemia viral (v-erb-b2) oncogene homolog 2	164870:Adenocarcinoma of lung, somatic, 211980 (3);Glioblastoma, somatic, 137800 (3);Gastric cancer, somatic, 613659 (3);Ovarian cancer, somatic, (3)
BRCA1	Breast cancer-1 gene	113705:{Breast-ovarian cancer, familial, 1}, 604370 (3);{Pancreatic cancer, susceptibility to, 4}, 614320 (3)
BRIP1	BRCA1-associated C-terminal helicase 1	605882:Breast cancer, early-onset, 114480 (3);Fanconi anemia, complementation group J, 609054 (3)
SMARCA4	SWI/SNF-related, matrix-associated, actin-dependent regulator of	603254:{Rhabdoid tumor predisposition syndrome 2}, 613325 (3);Mental retardation, autosomal dominant 16, 614609 (3)
AURKA	Aurora kinase A	603072:{Colon cancer, susceptibility to}, 114500 (3)
EP300	E1A-binding protein, 300kD	602700:Rubinstein-Taybi syndrome 2, 613684 (3);Colorectal cancer, somatic, 114500 (3)
MLH1	mutL, E. coli, homolog of, 1	120436:Colorectal cancer, hereditary nonpolyposis, type 2, 609310 (3);Mismatch repair cancer syndrome, 276300 (3);Muir-Torre

Gene	OMIM Title	OMIM Disorders
		syndrome, 158320 (3)
FGFR3	Fibroblast growth factor receptor-3	134934:Achondroplasia, 100800 (3);Hypochondroplasia, 146000 (3);Thanatophoric dysplasia, type I, 187600 (3);Crouzon syndrome with acanthosis nigricans, 612247 (3);Muenke syndrome, 602849 (3);Bladder cancer, somatic, 109800 (3);Colorectal cancer, somatic, 114500 (3);Cervical cancer, somatic, 603956 (3);LADD syndrome, 149730 (3);CATSHL syndrome, 610474 (3);Nevus, epidermal, somatic, 162900 (3);Thanatophoric dysplasia, type II, 187601 (3);Spermatocytic seminoma, somatic, 273300 (3)
APC	Adenomatous polyposis coli	611731:Adenomatous polyposis coli, 175100 (3);Gastric cancer, somatic, 613659 (3);Adenoma, periampullary, somatic (3);Hepatoblastoma, somatic, 114550 (3);Desmoid disease, hereditary, 135290 (3);Colorectal cancer, somatic, 114500 (3);Brain tumor-polyposis syndrome 2, 175100 (3);Gardner syndrome, 175100 (3)
EGFR	Epidermal growth factor receptor	131550:Nonsmall cell lung cancer, response to tyrosine kinase inhibitor in, 211980 (3);Adenocarcinoma of lung, response to tyrosine kinase inhibitor in, 211980 (3);{Nonsmall cell lung cancer, susceptibility to}, 211980 (3);?Inflammatory skin and bowel disease, neonatal, 2, 616069 (3)
BRAF	Murine sarcoma viral (v-raf) oncogene homolog B1	164757:Melanoma, malignant, somatic (3);Colorectal cancer, somatic (3);Adenocarcinoma of lung, somatic, 211980 (3);Nonsmall cell lung cancer, somatic (3);Cardiofaciocutaneous syndrome, 115150 (3);Noonan syndrome 7, 613706 (3);LEOPARD syndrome 3, 613707 (3)
PAX5	Paired box homeotic gene-5 (B-cell lineage specific activator	167414:{Leukemia, acute lymphoblastic, susceptibility to, 3}, 615545 (3)
PTCH1	Patched, Drosophila, homolog of	601309:Basal cell nevus syndrome, 109400 (3);Basal cell carcinoma, somatic, 605462 (3);Holoprosencephaly-7, 610828 (3)
FGFR2	Fibroblast growth factor receptor-2 (bacteria-expressed kinase)	176943:Crouzon syndrome, 123500 (3);Jackson-Weiss syndrome, 123150 (3);Beare-Stevenson cutis gyrata syndrome, 123790 (3);Pfeiffer syndrome, 101600 (3);Apert syndrome, 101200 (3);Saethre-Chotzen syndrome, 101400 (3);Craniosynostosis, nonspecific (3);Gastric cancer, somatic, 613659 (3);Craniofacial-skeletal-dermatologic dysplasia, 101600 (3);Antley-Bixler syndrome without genital anomalies or disordered steroidogenesis, 207410 (3);Scaphocephaly and Axenfeld-Rieger anomaly (3);LADD syndrome, 149730 (3);Scaphocephaly, maxillary retrusion, and mental retardation, 609579 (3);Bent bone dysplasia syndrome, 614592 (3)
WT1	Wilms tumor-1	607102:Wilms tumor, type 1, 194070 (3);Denys-Drash syndrome, 194080 (3);Nephrotic syndrome, type 4, 256370 (3);Frasier syndrome, 136680 (3);Meacham syndrome, 608978 (3);Mesothelioma, somatic, 156240 (3)
ATM	Ataxia-telangiectasia mutated (includes complementation groups A, C,	607585:Ataxia-telangiectasia, 208900 (3);Lymphoma, B-cell non-Hodgkin, somatic (3);{Breast cancer, susceptibility to}, 114480 (3);Lymphoma, mantle cell (3);T-cell prolymphocytic leukemia, somatic (3)
KMT2A	Lysine (K)-specific methyltransferase 2E	159555:Wiedemann-Steiner syndrome, 605130 (3);Leukemia, myeloid/lymphoid or mixed-lineage (2)
AKT1	Murine thymoma viral (v-akt) oncogene homolog-1	164730:Breast cancer, somatic, 114480 (3);Colorectal cancer, somatic, 114500 (3);Ovarian cancer, somatic, 167000 (3);{Schizophrenia, susceptibility to}, 181500 (2);Proteus syndrome, somatic, 176920 (3);Cowden syndrome 6, 615109 (3)
NF1	Neurofibromin (neurofibromatosis, type I)	613113:Neurofibromatosis, type 1, 162200 (3);Leukemia, juvenile myelomonocytic, 607785 (3);Neurofibromatosis, familial spinal, 162210 (3);Neurofibromatosis-Noonan syndrome, 601321 (3);Watson syndrome, 193520 (3)
SMAD4	Mothers against decapentaplegic, Drosophila, homolog of, 4	600993:Pancreatic cancer, somatic, 260350 (3);Juvenile polyposis/hereditary hemorrhagic telangiectasia syndrome, 175050 (3);Myhre syndrome, 139210 (3);Polyposis, juvenile intestinal, 174900 (3)
ALK	Anaplastic lymphoma kinase (Ki-1)	105590:{Neuroblastoma, susceptibility to, 3}, 613014 (3)
SRC	Protooncogene SRC, Rous sarcoma	190090:Colon cancer, advanced, somatic (3)

Gene	OMIM Title	OMIM Disorders
ATR	Anthrax toxin receptor 1;Ataxia-telangiectasia and Rad3-related (FRAP-related protein-1);Serpin peptidase inhibitor, clade A, member 2, pseudogene	606410:{Hemangioma, capillary infantile, susceptibility to}, 602089 (3);GAPO syndrome, 230740 (3);601215:Seckel syndrome 1, 210600 (3);Cutaneous telangiectasia and cancer syndrome, familial, 614564 (3);107410:None
ESR1	Estrogen receptor 1	133430:Estrogen resistance, 615363 (3);{HDL response to hormone replacement, augmented} (3);{Migraine, susceptibility to}, 157300 (3);{Atherosclerosis, susceptibility to} (3);{Myocardial infarction, susceptibility to}, 608446 (3);{Breast cancer}, 114480 (1)
MET	Oncogene MET;RNA guanine-7-methyltransferase	164860:Renal cell carcinoma, papillary, 1, familial and somatic, 605074 (3);Hepatocellular carcinoma, childhood type, 114550 (3);603514:None
JAK2	Janus kinase 2 (a protein-tyrosine kinase)	147796:Polycythemia vera, 263300 (3);Thrombocythemia 3, 614521 (3);Myelofibrosis, somatic, 254450 (3);{Budd-Chiari syndrome}, 600880 (3);Leukemia, acute myelogenous, 601626 (3);Erythrocytosis, somatic, 133100 (3)

7.3.1.3 Known EAC somatic alterations

Eight significantly mutated genes in EAC identified in the literature[206, 208] were analyzed as part of the pan-cancer panel, namely ARID1A, ARID2, CDKN2A (P16), KAT6A, PIK3CA, SMAD4, SMARCA4 and TP53. Of particular interest are TP53 and SMAD4, specific to only intervention-requiring disease (HGD and EAC) and not occurring in non-dysplastic BE epithelium. These mutations were examined in training cohort samples (for which we also have genome-wide methylation and expression profiling data), and summarized in Table 7-3 below.

Table 7-3: Known EAC somatic alterations (ARID1A, ARID2, CDKN2A (P16), KAT6A, PIK3CA, SMAD4, SMARCA4 and TP53) examined as part of the pan-cancer dataset. Samples were nine pairwise disease-normal analysis at all stages of esophageal adenocarcinoma development (n=2 non-dysplastic BE, n=3 mixed BE-LGD, n=2 LGD, n=2 HGD/EAC; each with matched normal squamous esophageal mucosa pairing). Results returned SNPs and indels classified as either high or medium impact by VEP (ordered by impact severity). No CDKN2A (P16) mutations were returned for these samples at the assigned level of impact severity.

Variant	Type	Gene	Impact	Impact severity	Amino acid change
chr1:g.27057682C>T	SNP	ARID1A	stop_gained	HIGH	Q/*
chr1:g.27106373AGGC	Deletion	ARID1A	frameshift_variant	HIGH	GNDFEMSKH

Variant	Type	Gene	Impact	Impact severity	Amino acid change
AATGACTTTGAGATGTC CAAACACCCAG>A					PG/X
chr1:g.27106861C>T	SNP	ARID1A	stop_gained	HIGH	R/*
chr12:g.46287298C>A	SNP	ARID2	stop_gained	HIGH	S/*
chr17:g.7578263G>A	SNP	TP53	stop_gained	HIGH	R/*
chr19:g.11105636G>T	SNP	SMARCA4	stop_gained	HIGH	E/*
chr12:g.46125065C>A	SNP	ARID2	missense_variant	MED	N/K
chr12:g.46215233A>G	SNP	ARID2	missense_variant	MED	E/G
chr12:g.46230697C>A	SNP	ARID2	missense_variant	MED	L/I
chr12:g.46243873C>A	SNP	ARID2	missense_variant	MED	A/D
chr12:g.46244790C>A	SNP	ARID2	missense_variant	MED	L/I
chr12:g.46244919C>G	SNP	ARID2	missense_variant	MED	Q/E
chr12:g.46244965C>A	SNP	ARID2	missense_variant	MED	A/E
chr12:g.46246024G>C	SNP	ARID2	missense_variant	MED	S/T
chr12:g.46254685G>T	SNP	ARID2	missense_variant	MED	M/I
chr17:g.7577034C>A	SNP	TP53	missense_variant	MED	G/W
chr17:g.7577105G>A	SNP	TP53	missense_variant	MED	P/L
chr17:g.7577121G>A	SNP	TP53	missense_variant	MED	R/C
chr17:g.7577575A>C	SNP	TP53	missense_variant	MED	Y/D
chr17:g.7578475G>C	SNP	TP53	missense_variant	MED	P/R
chr17:g.7578526C>A	SNP	TP53	missense_variant	MED	C/F
chr18:g.48591925G>T	SNP	SMAD4	missense_variant	MED	C/F
chr19:g.11101922C>T	SNP	SMARCA4	missense_variant	MED	R/C
chr19:g.11101935T>C	SNP	SMARCA4	missense_variant	MED	I/T
chr19:g.11144811G>C	SNP	SMARCA4	missense_variant	MED	V/L
chr19:g.11145812C>G	SNP	SMARCA4	splice_region_variant	MED	None
chr19:g.11170507G>T	SNP	SMARCA4	missense_variant	MED	G/C
chr3:g.178917565G>T	SNP	PIK3CA	missense_variant	MED	C/F
chr3:g.178927423C>T	SNP	PIK3CA	missense_variant	MED	L/F
chr3:g.178943747C>A	SNP	PIK3CA	splice_region_variant	MED	None
chr3:g.178952085A>G	SNP	PIK3CA	missense_variant	MED	H/R
chr8:g.41790102G>T	SNP	KAT6A	missense_variant	MED	P/Q
chr8:g.41790234G>T	SNP	KAT6A	missense_variant	MED	T/N
chr8:g.41790661G>T	SNP	KAT6A	missense_variant	MED	P/T
chr8:g.41791355C>A	SNP	KAT6A	missense_variant	MED	Q/H
chr8:g.41795055C>A	SNP	KAT6A	missense_variant	MED	R/L
chr8:g.41832273C>A	SNP	KAT6A	missense_variant	MED	M/I

As expected, the majority of (seven of eight) known genes mutated in EAC included in the pan-cancer screening panel were detected in our paired disease-normal analysis. These were typically single nucleotide polymorphisms, with multiple varied point mutations occurring within each of the genes examined. Missense and splice region variants (medium impact severity) were much more common than stop gain or frameshift variants (high impact severity).

The disease-classification distribution of known EAC somatic mutation was examined across our dataset (Figure 7-4). SMAD4 and TP53 mutation are reported to be specific to intervention-requiring disease whereas the remaining six known EAC mutations are also detectable in non-dysplastic

BE. Our data supported literature observations, with SMAD4 only detectable in HGD/EAC samples and TP53 in dysplastic disease and HGD/EAC, but not detected in non-dysplastic disease. Interestingly, in our samples, ARID1A was observed to be similar to TP53 in disease classification distribution and was also not detected in non-dysplastic BE. However this observation is from only a small number of samples and we hypothesize that with increased sample size it would also possibly be observed in non-dysplastic BE samples. All other genes had mutations that were detectable in non-dysplastic BE mucosa.

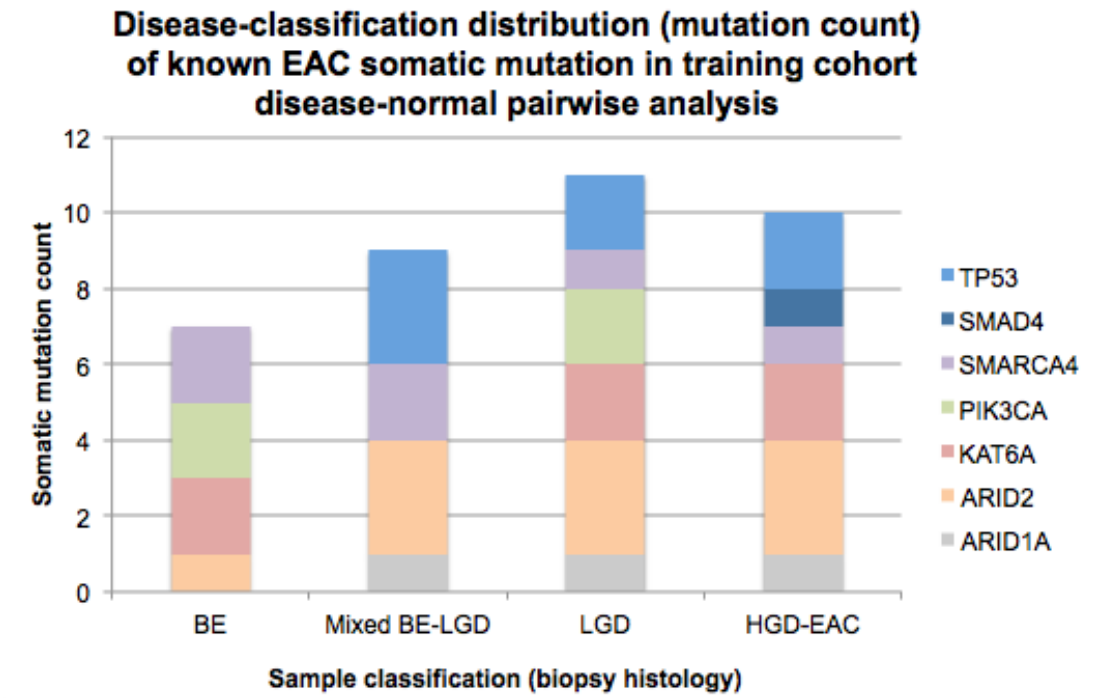


Figure 7-4: Disease-classification distribution of known EAC somatic mutation in training cohort disease-normal pairwise analysis. ARID1A, ARID2, KAT6A, PIK3CA, SMAD4, SMARCA4 and TP53 were evaluated for mutational count within each disease class (BE, mixed BE-LGD, LGD, HGD-EAC) in disease-normal pairwise analysis of training cohort samples. Both medium and high impact mutations were included.

7.3.2 CIMP causal mutation in the esophagus

CIMP in EAC has been reported as early as 2001, however analysis has typically focused on marker genes for colorectal CIMP, where it was first identified[47, 49], rather than establishing a specific esophageal adenocarcinoma CIMP profile (E-CIMP). In 2011, Kaz et al observed methylation sub-groups 'high methylation epigenotype' (HME) and 'low methylation epigenotype' (LME) in both BE and EAC, but did not attempt to attribute it to a causal genetic mutation or demographic/phenotypic feature[238]. Since then, the molecular basis of CIMP in EAC has still not been addressed. Our pan-cancer panel includes sequencing of IDH1, IDH2, TET2, KRAS and BRAF mutations, all of which have been previously implicated in CIMP formation in other cancers[214, 215, 229]. We examined all dysplastic and EAC tissues for evidence of any high or medium impact, disease-associated SNP or indels present (Figure 7-5) and found minimal evidence of mutation in these genes in the development of EAC. This supports previously reported findings that the molecular basis of CIMP in esophageal adenocarcinoma is very different to that in glioma or AML; and even varies from that of other solid GI tumors such as colorectal cancer.

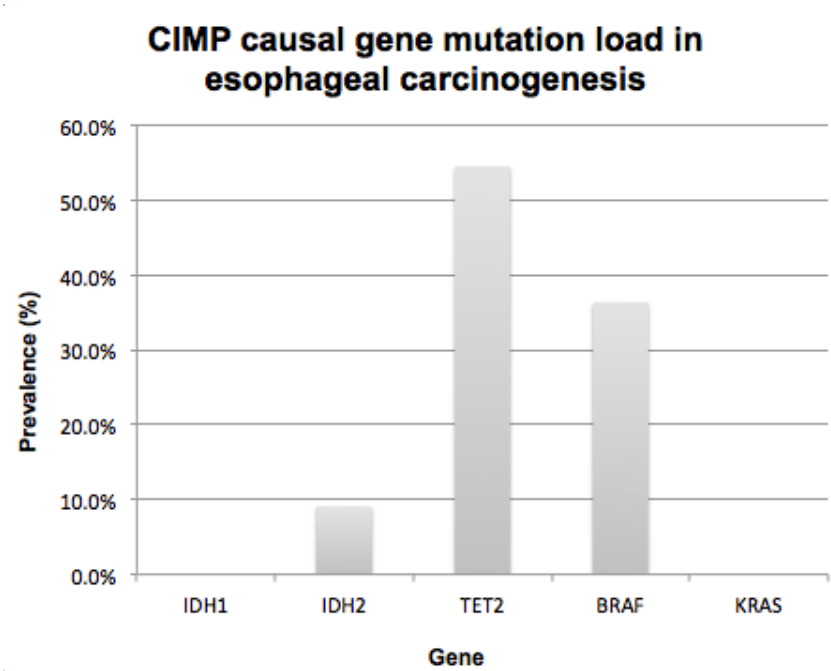


Figure 7-5: CIMP causal gene mutational load in esophageal carcinogenesis. Prevalence of cancer-associated somatic mutation in genes implicated in

CIMP causality in other cancers (IDH1, IDH2, TET2, BRAF and KRAS) in samples at all stages of esophageal adenocarcinoma development. Results expressed as percentage prevalence across all samples analyzed.

7.3.2.1 IDH1/2 and TET2 mutation in the esophagus

In matched disease-normal pairwise analysis, there was evidence of an IDH1 SNP (chr2:g.209116240C>G, missense variant, codon change gaG/gaC, amino acid change Glu/Asp) in a single LGD patient, notably NOT occurring at R132, as found in glioma[214]. This same patient also showed evidence of an IDH2 SNP (chr15:g.90631586C>A, splice region variant, no codon or amino acid change). Two TET2 mutations were detected in a single EAC patient (both missense variant, chr4:g.106157848C>A, chr4:g.106196343C>A, codon change Caa/Aaa, tCt/tAt, amino acid change Gln/Lys, Ser/Tyr respectively). There is no evidence for TET2 and IDH1/2 as mutually exclusive mutations in our data. These data support previously reported findings that the molecular basis of CIMP in solid tumors such as esophageal adenocarcinoma is very different to that of glioma or AML[212].

7.3.2.2 BRAF and KRAS mutation in the esophagus

There was evidence of four different BRAF single nucleotide polymorphisms (SNPs) in our pan-cancer genetic screening panel. Two of these mutations occurred in a single LGD patient, the other two mutations were evident in two further individuals. BRAF mutations are as outlined in Table 7-4.

Table 7-4: BRAF mutation occurring in esophageal tissue matched disease-normal pairwise analysis

Variant	Impact	Codon change	Amino acid change
chr7:g.140434499C>A	Missense variant	ttG/ttT	Leu/Phe
chr7:g.140439677C>A	Missense variant	Gcc/Tcc	Ala/Ser
chr7:g.140494197C>T	Missense variant	Gaa/Aaa	Glu/Lys
chr7:g.140501303G>C	Missense variant	Cag/Gag	Gln/Glu

We did not detect any KRAS mutations in our training cohort samples. This is somewhat surprising as both KRAS and BRAF are members of the ras signal transduction pathway and are typically mutually exclusive in colorectal cancer[229]. Lack of detection of KRAS mutation may be due to the low numbers of samples analysed (9 disease-normal pairs: BE (n=2), BE-LGD mixed (n=3), LGD (n=2), EAC-HGD (n=2)), further confounded by the variability in genetic signature commonly observed in EAC tissues[207].

7.4 Discussion

The pan-cancer mutational profiling panel has continued to develop as publications identify possible further targets. An additional 12 targets have been flagged for inclusion, namely AhR (aryl hydrocarbon receptor), SDHAF2[§], TMEM127[§], MAX[§], EGLN1[§], HIF2A[§], KIF1B[§], EPHB2, ARMC5, LTS2, PHD2, GR101. Of particular interest for esophageal adenocarcinoma is transmembrane protein 127 (TMEM127), a tumor susceptibility gene. Genes from this family (TMEM40 and TMEM109) are included on patent PCT/US2011/041659, Method for determining predisposition to esophageal and esophageal-related disorders (Esophageal disorders: esophagitis, squamous cell carcinoma or esophageal, cardia or gastric adenocarcinoma; Esophageal-related disorders: head and neck cancer, throat cancer, gastric cancer or mouth cancer). Furthermore, in a 2013 exome and whole-genome sequencing study, Dulak et al examined recurrent driver events and mutational complexity in esophageal adenocarcinoma, finding TMEM132D to be altered in 14 of 146 cases (9.6%), placing it in the top 30 mutated genes in EAC[208].

Somatic mutation pan-cancer profiling of known EAC alterations in the training cohort supports observations reported in literature that the majority of alterations occur early in the metaplasia-dysplasia-adenocarcinoma sequence. Weaver et al (2014) reported only TP53 and SMAD4 to be specific for HGD and EAC, and not present in non-dysplastic BE[206]. Our

[§] From Martins and Bugalho. 2014. Int J Endocrinol. Epub 2014 May 12[239].

findings supported this observation, however interestingly we note that TP53 mutation is also detected in low-grade dysplastic Barrett's mucosa. These observations (that the majority of known EAC somatic mutation are already detectable in non-dysplastic BE mucosa) support the hypothesis that EAC is not a mutation-driven cancer.

Aside from known EAC mutations, a large number of alterations across a wide-range of genes were present in not only EAC samples, but also low-grade dysplastic and non-dysplastic Barrett's mucosa. This is interesting as pan-cancer panel development has included genes far removed from those expected to be involved in the development of gastrointestinal neoplasia; including familial pituitary genes and pituitary development (embryonic) genes. We observed a very low rate of universal alteration in our samples. Very few genes had the same mutation detected in all diseased samples. More commonly, genes would have various point mutations across various samples. For example, twelve BRCA2 single nucleotide polymorphisms were observed across all but two of our disease-normal pairs (one non-dysplastic BE, one mixed BE-LGD), with one EAC sample containing 5 different BRCA2 mutations, but the majority of other samples each containing a single unique BRCA2 SNP.

With regard to somatic pan-cancer mutation load observed in our samples (representing the full spectrum of disease progression), we note a general trend in increased mutational load as disease progresses through the metaplasia-dysplasia-neoplasia sequence, with some anomalies, such as a very low mutational load (similar to non-dysplastic BE samples) in an advanced invasive EAC sample. This is not too surprising, given the evidence that EAC is not considered to be a mutation-driven cancer and the highly variable genomic signatures and wide range of mutation rates observed in of EAC tumors[49, 90, 206, 208].

A CpG island methylator phenotype (CIMP) phenotype has been observed in many cancers, however the molecular basis of this phenotype is not universal across all cancers. CIMP in glioma is reported to be established by,

and highly dependent on, IDH1 mutation, which alters histone marks and induces extensive DNA hypermethylation[214]. Similarly in acute myeloid leukemia, IDH1, IDH2 and TET2 mutations were found to induce a global hypermethylation signature[215]. However it has been reported that in many solid tumors such as gastrointestinal stromal tumors (GIST), bladder, breast, colorectal, lung, ovarian, pancreas, prostate and thyroid carcinoma[212, 216-218], IDH1 mutation is absent, despite CIMP being reported in these tumors[219-226]. The molecular basis of CIMP in colorectal cancer has been suggested to be either BRAF or KRAS mutation driven, however a causal relationship has not been established and the inverse, that CIMP may provide a favorable environment for mutation to occur, has been proposed[230]. No esophageal adenocarcinoma CIMP profile has been established, and despite variations in CIMP causation across cancer types, a colorectal CIMP profile has continued to be used to classify EAC-CIMP as recently as this year[49]. We investigated levels of IDH1, IDH2, TET2, KRAS and BRAF mutation in our esophageal adenocarcinoma and developmental samples and found low mutational load, supporting the hypothesis that alterations in these genes are not causal for EAC and more work is needed to define an EAC CIMP profile and determine a causal factor for this phenotype.

CHAPTER 8: GENERAL DISCUSSION AND CONCLUDING REMARKS

8.1 Introduction

In this thesis, I have addressed the problem of poor esophageal adenocarcinoma outcome due to lack of an early stage diagnostic test, by identification and validation of aberrant methylation, an early and frequent event in carcinogenesis, as a biomarker for early detection of esophageal adenocarcinoma and dysplastic forms of its precursor disease, Barrett's esophagus.

In this final chapter I will summarize major experimental findings and put these into context of the current field of methylation and esophageal adenocarcinoma diagnostics.

8.2 Overview of experimental findings

The major findings of this thesis are:

1. **RNA preservation solutions work effectively at maintaining isolated RNA integrity in esophageal tissue, but introduce problems associated with high salt concentration.** Esophageal tissue samples were placed directly into RNA*later*, which fully penetrates the tissue and preserves RNA for downstream applications, but introduces purity problems due to high salt concentration. Special care needs to be taken extracting nucleic acids to ensure high yield and purity.
2. **Current 'gold standard' endoscopic/histologic methods are inadequate for robust, consistent diagnosis of low-grade dysplasia.** Low-grade dysplastic samples are of exceptional importance as they represent early stage carcinogenic change, but diagnostic difficulties may introduce bias to potential biomarker identification. To address this, low-grade dysplastic tissues were

included at every stage of the study, but not used for bioinformatic analyses to determine potential biomarker candidates, thereby not able to introduce error due incorrect diagnosis, but able to be checked to determine if selected biomarker candidates were useful in these very early stage samples. Sample heterogeneity cut-offs were applied to ensure robust, reliable determination of biomarker candidates.

3. **Comprehensive same sample genome-wide methylation and expression profiling from samples at every stage of the metaplasia-dysplasia-adenocarcinoma sequence is novel for this disease.** Disease-associated differential methylation exists and is detectable between non-dysplastic BE and BE with areas of low-grade dysplasia. This is clinically important for early-stage detection biomarkers. The use of non-dysplastic BE as baseline compared to combined HGD-EAC was informative for methylation changes associated with early stage carcinogenic development.
4. **The most significantly hypermethylated regions in HGD and EAC (with respect to normal esophageal epithelium) were, in the most part, also significantly hypermethylated in non-dysplastic BE, but unmethylated in normal peripheral blood as well as normal esophageal, duodenal and proximal stomach mucosa.** This highlights problems associated with simple normal versus cancer analyses for esophageal adenocarcinoma biomarker identification, as these regions are not specific to early-stage carcinogenesis and cancer, but also identify common, benign precursor disease. Furthermore, the most significantly hypermethylated regions in BE with respect to normal esophageal epithelium are also significantly hypermethylated in HGD and EAC, but unmethylated in normal peripheral blood as well as normal esophageal, duodenal and proximal stomach mucosa. Thus, we can infer that disease-associated methylation change occurs early, and is maintained throughout the metaplasia-dysplasia-adenocarcinoma progression.
5. **Much of the hypomethylation occurring in the metaplasia-dysplasia-adenocarcinoma sequence is attributable to maintenance of columnar mucosa rather than disease-associated**

progression through the development sequence. For this reason, no hypomethylation targets were carried through to validation and targeted sequencing studies.

6. **Much of the significant differential expression (both up- and down-regulated) is also attributable to the maintenance of columnar mucosa rather than disease progression through the metaplasia-dysplasia-adenocarcinoma sequence.** Thus we can conclude that tissue type difference (squamous cells comprising normal, healthy esophageal mucosa, as opposed to columnar mucosa present in BE, HGD, EAC and control tissues) has a more significant impact on global expression profile than disease development.
7. **Esophageal adenocarcinoma is a complex disease with many different molecular subtypes, leading to large variation in global expression patterns between individuals.** Significantly less intra-classification variation was observed in global methylation than expression, supporting methylation biomarkers as more viable than expression biomarkers for universally detectable change irrespective of variable esophageal adenocarcinoma subtypes.
8. **Methylation-expression correlation analyses support recent literature claiming more complex interactions in cancer development than simplistic 'hypomethylation resulting in transcriptional activation and hypermethylation triggering transcriptional silencing'.** Whilst these associations were observed in our data sets, they were by no means the predominant occurrence. Disease-associated differential methylation at or near the transcription start site resulted in statistically significant ($p \leq 0.05$) alteration in expression of these genes in just 18% of target regions selected for validation.
9. **Gene set enrichment revealed expected functional association as disease developed as well as highlighting the importance of precursor versus intervention-requiring disease for enrichment of disease associated change.** Enriched gene sets in esophageal adenocarcinoma development include those for angiogenesis, inflammatory response, cytokine (interferon alpha and interferon

gamma) response as well as p53 pathways. Gene sets showing early enrichment in precursor disease as well as adenocarcinoma include KRAS signaling, epithelial-mesenchymal transition and coagulation. Cancer neighborhood gene sets selected as the most enriched using N v HGD-EAC showed significantly increased fold enrichment and gene count in the BE v HGD-EAC comparison.

10. **Increased methylation of tumor suppressor p63 promoter region may be a contributing factor to reduced p63 expression and possibly implicated in BE development and/or carcinogenic progression** Our data support disease-associated tumor suppressor p63 promoter methylation in one of two p63 promoter regions in BE, HGD and EAC samples compared to normal and gastro-esophageal reflux disease samples. Loss of tumor suppressor p63 expression is associated with carcinogenesis in a number of cancer types. Its importance in the origin of BE is highlighted in experiments showing that a p63 null mouse quickly develops Barrett's-like tissue.
11. **Adaptation of sequencing protocols was successful for short amplicons in tissue samples with direct transferability to blood biomarker investigation.** Bisulfite-specific PCR for unbiased amplification of specific regions of interest combined with targeted amplicon sequencing was used to validate potential methylation biomarker candidates. Assay design and protocol adaptation were performed so that techniques are immediately transferable to the short, fragmented circulating DNA extracted from blood.
12. **The use of control (proximal stomach and duodenal) tissues with similar mucosal structure to Barrett's esophageal tissue enable effective filtering of aberrant methylation indicative of maintenance of columnar epithelium.** Control normal peripheral blood was used to further filter biomarker candidates to ensure suitability for future blood test. The resultant subset of disease-associated aberrant methylation is clinically applicable for an early cancer detection test.
13. **Potential hypermethylation biomarkers include ARL10, a region upstream of CA4, CACNA2D2, KLF7, LRRC43, TEPP, TUBA3FP,**

VANGL2, ZNF221 and ZNF699. Disease-associated hypermethylated target regions indicative of early esophageal adenocarcinoma were established using stringent cut-off criteria applied to global methylation profiling data (little or no methylation in non-dysplastic BE, normal esophageal mucosa and control tissues as well as normal peripheral blood) and validated using BSP amplification and target amplicon sequencing. Global methylation profiling (n=250 samples) of an external data set (the only published HM450 data set for esophageal adenocarcinoma) were used for further validation. A number of these candidates are novel, not only for esophageal adenocarcinoma identification, but for their role in carcinogenesis (for example ARL10 and TUB3FP); others, such as KLF7 and VANGL2 have family members associated with carcinogenesis in other cancer types, but are novel for EAC carcinogenesis.

14. **Aberrant promoter methylation of DNA repair gene MGMT plays a significant role in the multistep process of Barrett's carcinogenesis.** Our data do not support literature surrounding the reported diagnostic utility of MGMT promoter methylation in esophageal adenocarcinoma, due to methylation observed in a subset of non-dysplastic BE. However the very low methylation levels observed in normal squamous epithelia, GERD and stomach tissues do support a disease-associated role. Early epigenetic silencing of DNA repair gene MGMT (at the non-dysplastic BE stage) would likely increase mutation rates, one or more which may provide the cell with selective advantage. Carried on, continued epigenetic repression of MGMT would generate further mutations, which may produce tumor. We can hypothesize from our data that the subset of non-dysplastic BE patients showing promoter MGMT hypermethylation may be more susceptible to malignant change. However it would not be possible to predict disease progression using this information alone. It is likely that epigenetic MGMT repression would need to occur in co-ordination with repression of other DNA repair genes or other biological events.
15. **Bimodal methylation clustering (hypermethylation of a subset of a particular disease class, with another subset entirely**

unmethylated) was observed across all stages of Barrett's carcinogenesis. We hypothesize that this may be, in part, due to the inflammation-triggered stem cell origin of BE resulting in molecularly disparate groups of disease arising, as observed in expression profiling of all stages of disease development. Chromothripsis, massive genomic rearrangement more prevalent in EAC then any other cancer, may also contribute to the bimodality in EAC samples.

16. **Novel potentially clinical useful methylation biomarkers have been identified as a result of this study**. A triplex of hypermethylation in any of TUBA3FP, VANGL2 or ARL10 was able to accurately identify all HGD and EAC patients with 100% sensitivity (beta value cut-offs 0.05, 0.02 and 0.08 respectively) with 84.6% specificity (based on methylation detected in patients with a normal, healthy esophagus, or non-dysplastic BE). The triplex was able to successfully identify 95% of all LGD, HGD and EAC patients (specificity 76.9%).
17. **There is evidence of application of identified biomarkers for predicting dysplastic progression.** TEPP and a target region upstream of CA4 showed significant hypermethylation in a non-dysplastic BE sample taken one year *before* development of low-grade dysplasia. Aberrant methylation of these target regions may indicate predisposition to dysplastic progression.
18. **There is evidence of application of identified biomarkers for predicting the necessity of treatment for low-grade dysplasia.** Treatment of low-grade dysplasia is still disputed and discussed in guidelines for the management of Barrett's worldwide. The ability to predict progression from low to high-grade dysplasia may aid in determining how aggressively to treat LGD. Methylation of LRRC43 and ZNF699 in an LGD patient was consistent with a more advanced disease profile 8 months prior to HGD progression, despite several endoscopic mucosal resection treatments. Similarly, biomarkers for the natural regression of LGD to non-dysplastic BE may aid in treatment determination and/or surveillance regime for LGD patients. For example, a patient with persistent long segment LGD (untreated,

constant surveillance) was observed to have complete absence of methylation in a number of target regions, including ARL10, TUBA3FP, CACNA2D2, LRRC43 and TRANK1. On follow-up, this patient was diagnosed as being without dysplasia two years after the LGD biopsy was taken and analyzed for this study.

19. **There is evidence of application of identified biomarkers for disease monitoring with treatment.** A HGD patient was monitored and successfully treated over three visits in the space of 7½ months. Diagnoses of HGD (visit 1), LGD (visit 2) and non-dysplastic BE, no evidence of dysplasia (visit 3) were reflected in the methylation profile of TNFAIP8L3 and ZNF699, which showed significant correlation with disease regression as the patient was treated. Methylation of these target regions may be clinically valuable for monitoring disease status with treatment and have possible application in predicting disease recurrence as part of a screening panel.
20. **Increased mutational load was detected as disease progressed through the metaplasia-dysplasia-adenocarcinoma sequence.** Interesting was the detection of cancer-associated mutational load in non-dysplastic Barrett's mucosa, a benign epithelium. This is in line with observations that disease-associated methylation change occurs early, and is maintained throughout the metaplasia-dysplasia-adenocarcinoma progression. An interesting anomaly was an advanced, invasive carcinoma sample showing a distinct lack of cancer-associated somatic mutation; again highlighting the variation in molecular subtypes of esophageal adenocarcinoma.
21. **Our data supported current literature regarding somatic mutation in EAC.** Pairwise disease-normal analyses showed TP53, SMAD4 and ARID1A mutations to be specific to intervention-requiring disease and not detectable in non-dysplastic BE epithelium. Mutations in other genes such as SMARCA4, PIK3CA, KAT6A and ARID2 were observed in non-dysplastic BE as well as more advanced disease.
22. **There is minimal evidence to support esophageal CIMP causal mutation in genes implicated in CIMP formation in other cancers.** IDH1, IDH2, TET2, KRAS and BRAF mutation have been implicated in

CIMP formation in other cancers (such as glioma, AML and colorectal cancer) but, based on our profiling, do not show evidence of being CIMP causal mutations in esophageal adenocarcinoma.

8.3 Findings in context and future aspects

Since the commencement of my thesis, the first study using HM450 technology applied to esophageal adenocarcinoma tissues was only recently published by Krause et al[49] in February 2016. The group kindly agreed to share their data for use as my external validation data set prior to publication. Their study examines methylation profiles of normal, BE and EAC samples, but does not address intermediate dysplastic stages of carcinogenic development. To date, my study is the first to perform whole genome methylation profiling using comprehensive HM450 arrays on both low and high-grade dysplastic Barrett's tissue.

Current endoscopic diagnosis methods for EAC and its precursor disease, BE, are not viable for population screening, contributing to late stage EAC detection and poor prognosis. Considerable difficulties exist with current endoscopic/histologic diagnosis methods, especially surrounding correct diagnoses for intermediate disease development stages. A number of issues such as tissue heterogeneity, pathologist interpretation, sampling error and histology processing artifacts contribute to variation in diagnosis. We worked around this problem by exclusion of data pertaining to LGD samples for biomarker identification, but checking performance of selected biomarkers in these important samples indicative of early stage carcinogenic transformation. An alternative, assumption-free approach would be to use molecular signature (methylation, expression, mutational profile) to group patients, blinded to patient/sample diagnosis as defined by endoscopy/histology. We considered this alternative approach, however it quickly became complicated due to the molecularly disparate forms of EAC and the resultant large numbers of molecularly distinct groups; however this approach may be useful in cancer types with more uniform molecular disease profiles.

Targeted amplicon sequencing (used for validation of methylation profiling in this study) has been used successfully monitor TP53 mutation in ovarian cancer patients with sensitivity and specificity >97%[71]. The BSP assays for targeted amplicon sequencing used as part of this study were designed with application as a blood test in mind: a shorter amplicon length than usually used for amplicon sequencing is appropriate for amplification of cell-free circulating (cfc)-DNA extracted from blood.

Using (i) comprehensive methylation HM450 profiling, (ii) high-grade dysplasia included with EAC for all analyses, (iii) validation of selected regions in low-grade dysplasia, (iv) non-dysplastic BE as baseline and intervention-requiring disease (HGD-EAC) as the comparison group and (v) comprehensive controls including normal squamous epithelia, proximal stomach / duodenal tissues (columnar mucosal structure similar to that found in BE), and peripheral blood from normal healthy patients, enriched results for disease-associated methylation change indicative of early phase EAC development and *enabled identification of novel biomarkers* with potential application as a screening blood test. Biomarkers for esophageal adenocarcinoma to date have been limited in their clinical application, however my rigorous approach to biomarker identification, combined with new technology may help turn this around.

The identification of novel aberrant DNA methylation in EAC carcinogenesis has far-reaching implications not only as biomarkers for early identification, but also for cancer treatment, as recently shown in a landmark paper by Rivenbark et al[240], epigenetically reprogramming breast cancer cells by targeting DNA methylation.

The focus for this study is identification of potentially clinically useful methylation biomarkers for identification of early phase esophageal adenocarcinoma. The study was designed with a view to blood investigation (all target regions are essentially unmethylated in peripheral blood from healthy patients and amplification assays for targeted sequencing are

suitable for degraded, shorter fragments of cell-free circulating DNA in blood), which, if successful, would be applicable for screening of high-risk populations: middle-aged Caucasian males or individuals with a family history. A blood test would allow for easily implemented, cost-effective screening of the high-risk population, enabling detection of EAC and HGD at a curable stage, and ultimately providing an opportunity to improve EAC incidence, prognosis and mortality.

In the case of cancer patients, isolated cfc-DNA is known to reflect epigenetic characteristics of tumor occurring elsewhere in the body. In a 2014 study, Bettegowda et al were able to detect cfc-DNA with tumor origin from the blood of >75% of patients with advanced pancreatic, ovarian, colorectal, bladder, gastroesophageal, breast, melanoma, hepatocellular, and head and neck cancers[241]. The huge potential of 'liquid biopsy' (advantageous as it is widely accepted, readily repeated, convenient, non-invasive and low cost, also overcoming heterogeneity issues associated with tissue sampling) has been the focus for Dawson et al, analysing cfc-DNA in blood to assess tumour burden and treatment response in real time breast cancer management[242]. In a single-case study of serial plasma samples from a metastatic ER-positive and HER2-positive breast cancer patient, receiving two lines of targeted therapy over three years, Murtaza et al showed mutation levels in plasma reflect clonal hierarchy inferred from sequencing of tissue tumour biopsies, and plasma level of mutations specific to certain metastases correlate with treatment responses at these sites[72]. These studies highlight how cfc-DNA from plasma is a valuable, easily-accessible way to monitor or screen for cancerous changes occurring in tissues.

Additionally, this study has formed the foundation for further work (and subsequent funding applications) into the investigation of disease progression and regression. Further analysis of subsequent samples taken from patients analysed here that have progressed (or regressed) in their disease classification is currently underway to confirm hypotheses from this this and to further investigate clinical utility of these biomarkers.

8.4 Concluding remarks

The aims of this thesis were (i) to perform genome-wide methylation and expression profiling at all stages of the metaplasia-dysplasia-adenocarcinoma sequence to uncover novel disease-associated aberrant methylation with potential clinical application, (ii) comprehensively validate a subset of potential methylation biomarkers, (iii) evaluate performance of validated biomarkers with regard to diagnostic utility and other clinical application and (iv) examine cancer-associated mutational load in the same samples used for expression and methylation profiling to gain a more complete understanding of molecular change occurring in the development of esophageal adenocarcinoma.

By using well-classified patient data and stringent, quality controlled biospecimen selection for training and validation cohorts, I uncovered regions of disease-associated aberrant methylation novel for esophageal carcinogenesis. Utilizing comprehensive technical and independent validation by targeted amplicon sequencing and whole genome methylation profiling of a large external validation cohort, I demonstrated potential utility of these target regions for identification of intervention requiring disease. I propose a panel of three methylation biomarkers (TUBA3FP, VANGL2, ARL10) for identification of intervention requiring disease, reporting 100% sensitivity and 84.6% specificity and demonstrated biomarker application for prediction of disease progression as well as utility for monitoring disease status with treatment and predicting the necessity of treatment for low-grade dysplasia. By performing genome-wide methylation and expression profiling, as well as cancer-associated mutation screening on single tissue biopsies from all stages of the metaplasia-dysplasia-adenocarcinoma sequence, I was able to gain a more complete understanding of molecular change occurring in esophageal carcinogenesis.

Overall, the aims of this thesis have been met and I have been able to propose potentially clinically valuable methylation biomarkers for the detection of intervention-requiring disease, with potential application for non-

invasive, high-risk population screening for identification of esophageal adenocarcinoma at an early, treatable stage. The applicability of these novel epigenetic biomarkers to detect circulating tumor cells in blood to aid in clinical management will only be realized with further investigations conducted by our laboratory and clinical collaborators.

REFERENCES

1. Mikeska T, Bock C, Do H, Dobrovic A. DNA methylation biomarkers in cancer: progress towards clinical implementation. Expert review of molecular diagnostics. 2012;12(5):473-87.
2. Zhang J, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X, Ulyanov A, Wu G, Wilson M, Wang J, Brennan R, Rusch M, Manning AL, Ma J, Easton J, Shurtleff S, Mullighan C, Pounds S, Mukatira S, Gupta P, Neale G, Zhao D, Lu C, Fulton RS, Fulton LL, Hong X, Dooling DJ, Ochoa K, Naeve C, Dyson NJ, Mardis ER, Bahrami A, Ellison D, Wilson RK, Downing JR, Dyer MA. A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature. 2012;481(7381):329-34.
3. McKenna ES, Sansam CG, Cho YJ, Greulich H, Evans JA, Thom CS, Moreau LA, Biegel JA, Pomeroy SL, Roberts CW. Loss of the epigenetic tumor suppressor SNF5 leads to cancer without genomic instability. Molecular and cellular biology. 2008;28(20):6223-33.
4. Kanai Y. Genome-wide DNA methylation profiles in precancerous conditions and cancers. Cancer science. 2010;101(1):36-45.
5. Zhang Y. Epidemiology of esophageal cancer. World journal of gastroenterology : WJG. 2013;19(34):5598-606.
6. Behrens A, Ell C. [Challenges and limits for endoscopic resection of oesophageal and oesophagogastric cancer]. Zentralblatt fur Chirurgie. 2014;139(1):28-31.
7. Pouw RE, Gondrie JJ, Sondermeijer CM, ten Kate FJ, van Gulik TM, Krishnadath KK, Fockens P, Weusten BL, Bergman JJ. Eradication of Barrett esophagus with early neoplasia by radiofrequency ablation, with or without endoscopic resection. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 2008;12(10):1627-36; discussion 36-7.
8. Kanwal R, Gupta K, Gupta S. Cancer epigenetics: an introduction. Methods in molecular biology (Clifton, NJ). 2015;1238:3-25.
9. Feng Q, Yu M, Kiviat NB. Molecular biomarkers for cancer detection in blood and bodily fluids. Crit Rev Clin Lab Sci. 2006;43(5-6):497-560.
10. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, Clark SJ, Molloy PL. De novo identification of differentially methylated regions in the human genome. Epigenetics & chromatin. 2015;8:6.
11. Levenson VV, Melnikov AA. DNA Methylation as Clinically Useful Biomarkers—Light at the End of the Tunnel. Pharmaceuticals. 2012;5(12):94-113.
12. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nature reviews Genetics. 2002;3(6):415-28.
13. Kim M, Lee KT, Jang HR, Kim JH, Noh SM, Song KS, Cho JS, Jeong HY, Kim SY, Yoo HS, Kim YS. Epigenetic down-regulation and suppressive role of DCBLD2 in gastric cancer cell proliferation and invasion. Molecular cancer research : MCR. 2008;6(2):222-30.
14. Almeida M, Costa VL, Costa NR, Ramalho-Carvalho J, Baptista T, Ribeiro FR, Paulo P, Teixeira MR, Oliveira J, Lothe RA, Lind GE, Henrique R, Jeronimo C. Epigenetic regulation of EFEMP1 in prostate cancer:

biological relevance and clinical potential. Journal of cellular and molecular medicine. 2014;18(11):2287-97.
15. Bhat S, Kabekkodu SP, Noronha A, Satyamoorthy K. Biological implications and therapeutic significance of DNA methylation regulated genes in cervical cancer. Biochimie. 2016;121:298-311.
16. Diaz-Lagares A, Mendez-Gonzalez J, Hervas D, Saigi M, Pajares MJ, Garcia D, Crujeiras AB, Pio R, Montuenga LM, Zulueta J, Nadal E, Rosell A, Esteller M, Sandoval-Del Amor J. A novel epigenetic signature for early diagnosis in lung cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2016.
17. Jiang W, Cai R, Chen QQ. DNA Methylation Biomarkers for Nasopharyngeal Carcinoma: Diagnostic and Prognostic Tools. Asian Pacific journal of cancer prevention : APJCP. 2015;16(18):8059-65.
18. Margolin G, Petrykowska HM, Jameel N, Bell DW, Young AC, Elnitski L. Robust Detection of DNA Hypermethylation of ZNF154 as a Pan-Cancer Locus with in Silico Modeling for Blood-Based Diagnostic Development. The Journal of molecular diagnostics : JMD. 2016;18(2):283-98.
19. Salvianti F, Orlando C, Massi D, De Giorgi V, Grazzini M, Pazzagli M, Pinzani P. Tumor-Related Methylated Cell-Free DNA and Circulating Tumor Cells in Melanoma. Frontiers in molecular biosciences. 2015;2:76.
20. Sepulveda JL, Gutierrez-Pajares JL, Luna A, Yao Y, Tobias JW, Thomas S, Woo Y, Giorgi F, Komissarova EV, Califano A, Wang TC, Sepulveda AR. High-definition CpG methylation of novel genes in gastric carcinogenesis identified by next-generation sequencing. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2016;29(2):182-93.
21. Yoruker EE, Holdenrieder S, Gezer U. Blood-based biomarkers for diagnosis, prognosis and treatment of colorectal cancer. Clinica chimica acta; international journal of clinical chemistry. 2016;455:26-32.
22. Fukushige S, Horii A. DNA methylation in cancer: a gene silencing mechanism and the clinical potential of its biomarkers. The Tohoku journal of experimental medicine. 2013;229(3):173-85.
23. deVos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J, Steiger KV, Grutzmann R, Pilarsky C, Habermann JK, Fleshner PR, Oubre BM, Day R, Sledziewski AZ, Lofton-Day C. Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. Clinical chemistry. 2009;55(7):1337-46.
24. Gilbert MR, Dignam JJ, Armstrong TS, Wefel JS, Blumenthal DT, Vogelbaum MA, Colman H, Chakravarti A, Pugh S, Won M, Jeraj R, Brown PD, Jaeckle KA, Schiff D, Stieber VW, Brachman DG, Werner-Wasik M, Tremont-Lukats IW, Sulman EP, Aldape KD, Curran WJJ, Mehta MP. A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma. New England Journal of Medicine. 2014;370(8):699-708.
25. Schmidt B, Beyer J, Dietrich D, Bork I, Liebenberg V, Fleischhacker M. Quantification of cell-free mSHOX2 Plasma DNA for therapy monitoring in advanced stage non-small cell (NSCLC) and small-cell lung cancer (SCLC) patients. PloS one. 2015;10(2):e0118195.
26. Konecny M, Markus J, Waczulikova I, Dolesova L, Kozlova R, Repiska V, Novosadova H, Majer I. The value of SHOX2 methylation test in peripheral

blood samples used for the differential diagnosis of lung cancer and other lung disorders. Neoplasma. 2016;63(2).
27. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM. Multitarget stool DNA testing for colorectal-cancer screening. The New England journal of medicine. 2014;370(14):1287-97.
28. Sovich JL, Sartor Z, Misra S. Developments in Screening Tests and Strategies for Colorectal Cancer. BioMed Research International. 2015;2015.
29. Ahlquist DA, Zou H, Domanico M, Mahoney DW, Yab TC, Taylor WR, Butz ML, Thibodeau SN, Rabeneck L, Paszat LF, Kinzler KW, Vogelstein B, Bjerregaard NC, Laurberg S, Sorensen HT, Berger BM, Lidgard GP. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. Gastroenterology. 2012;142(2):248-56; quiz e25-6.
30. Lidgard GP, Domanico MJ, Bruinsma JJ, Light J, Gagrat ZD, Oldham-Haltom RL, Fourrier KD, Allawi H, Yab TC, Taylor WR, Simonson JA, Devens M, Heigh RI, Ahlquist DA, Berger BM. Clinical performance of an automated stool DNA assay for detection of colorectal neoplasia. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2013;11(10):1313-8.
31. Ahlquist DA, Skoletsky JE, Boynton KA, Harrington JJ, Mahoney DW, Pierceall WE, Thibodeau SN, Shuber AP. Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay panel. Gastroenterology. 2000;119(5):1219-27.
32. Delpu Y, Cordelier P, Cho WC, Torrisani J. DNA methylation and cancer diagnosis. International journal of molecular sciences. 2013;14(7):15029-58.
33. Phillips WA, Lord RV, Nancarrow DJ, Watson DI, Whiteman DC. Barrett's esophagus. Journal of gastroenterology and hepatology. 2011;26(4):639-48.
34. Ronkainen J, Aro P, Storskrubb T, Johansson SE, Lind T, Bolling-Sternevald E, Vieth M, Stolte M, Talley NJ, Agreus L. Prevalence of Barrett's esophagus in the general population: an endoscopic study. Gastroenterology. 2005;129(6):1825-31.
35. Sharma P, Dent J, Armstrong D, Bergman JJ, Gossner L, Hoshihara Y, Jankowski JA, Junghard O, Lundell L, Tytgat GN, Vieth M. The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. Gastroenterology. 2006;131(5):1392-9.
36. Sato F, Jin Z, Schulmann K, Wang J, Greenwald BD, Ito T, Kan T, Hamilton JP, Yang J, Paun B, David S, Olaru A, Cheng Y, Mori Y, Abraham JM, Yfantis HG, Wu TT, Fredericksen MB, Wang KK, Canto M, Romero Y, Feng Z, Meltzer SJ. Three-tiered risk stratification model to predict progression in Barrett's esophagus using epigenetic and clinical features. PloS one. 2008;3(4):e1890.
37. Bansal A, Lee IH, Hong X, Anand V, Mathur SC, Gaddam S, Rastogi A, Wani SB, Gupta N, Visvanathan M, Sharma P, Christenson LK. Feasibility of mcroRNAs as biomarkers for Barrett's Esophagus progression: a pilot cross-sectional, phase 2 biomarker study. Am J Gastroenterol. 2011;106(6):1055-63.

38. Jin Z, Cheng Y, Gu W, Zheng Y, Sato F, Mori Y, Olaru AV, Paun BC, Yang J, Kan T, Ito T, Hamilton JP, Selaru FM, Agarwal R, David S, Abraham JM, Wolfsen HC, Wallace MB, Shaheen NJ, Washington K, Wang J, Canto MI, Bhattacharyya A, Nelson MA, Wagner PD, Romero Y, Wang KK, Feng Z, Sampliner RE, Meltzer SJ. A Multicenter, Double-Blinded Validation Study of Methylation Biomarkers for Progression Prediction in Barrett's Esophagus. Cancer research. 2009;69(10):4112-5.
39. Clemons NJ, Phillips WA, Lord RV. Signaling pathways in the molecular pathogenesis of adenocarcinomas of the esophagus and gastroesophageal junction. Cancer biology & therapy. 2013;14(9):782-95.
40. Fels Elliott DR, Fitzgerald RC. Molecular markers for Barrett's esophagus and its progression to cancer. Curr Opin Gastroenterol. 2013;29(4):437-45.
41. Eloubeidi MA, Mason AC, Desmond RA, El-Serag HB. Temporal trends (1973-1997) in survival of patients with esophageal adenocarcinoma in the United States: a glimmer of hope? Am J Gastroenterol. 2003;98(7):1627-33.
42. Stavrou EP, McElroy HJ, Baker DF, Smith G, Bishop JF. Adenocarcinoma of the oesophagus: incidence and survival rates in New South Wales, 1972-2005. The Medical journal of Australia. 2009;191(6):310-4.
43. Reid BJ, Li X, Galipeau PC, Vaughan TL. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nature reviews Cancer. 2010;10(2):87-101.
44. Falk GW. Barrett's esophagus. Gastroenterology. 2002;122(6):1569-91.
45. Herman JG, Baylin SB. Gene Silencing in Cancer in Association with Promoter Hypermethylation. New England Journal of Medicine. 2003;349(21):2042-54.
46. Barton CA, Hacker NF, Clark SJ, O'Brien PM. DNA methylation changes in ovarian cancer: implications for early diagnosis, prognosis and treatment. Gynecologic oncology. 2008;109(1):129-39.
47. Eads CA, Lord RV, Wickramasinghe K, Long TI, Kurumboor SK, Bernstein L, Peters JH, DeMeester SR, DeMeester TR, Skinner KA, Laird PW. Epigenetic patterns in the progression of esophageal adenocarcinoma. Cancer research. 2001;61(8):3410-8.
48. Agarwal A, Polineni R, Hussein Z, Vigoda I, Bhagat TD, Bhattacharyya S, Maitra A, Verma A. Role of epigenetic alterations in the pathogenesis of Barrett's esophagus and esophageal adenocarcinoma. International journal of clinical and experimental pathology. 2012;5(5):382-96.
49. Krause L, Nones K, Loffler KA, Nancarrow D, Oey H, Tang YH, Wayte NJ, Patch AM, Patel K, Brosda S, Manning S, Lampe G, Clouston A, Thomas J, Stoye J, Hussey DJ, Watson DI, Lord RV, Phillips WA, Gotley D, Smithers BM, Whiteman DC, Hayward NK, Grimmond SM, Waddell N, Barbour AP. Identification of the CIMP-like subtype and aberrant methylation of members of the chromosomal segregation and spindle assembly pathways in esophageal adenocarcinoma. Carcinogenesis. 2016.
50. Clement G, Braunschweig R, Pasquier N, Bosman FT, Benhattar J. Methylation of APC, TIMP3, and TERT: a new predictive marker to

distinguish Barrett's oesophagus patients at risk for malignant transformation. The Journal of pathology. 2006;208(1):100-7.
51. Cancer Council Australia Barrett's Oesophagus Guidelines Working Party. Clinical practice guidelines for the diagnosis and management of Barrett's Oesophagus and Early Oesophageal Adenocarcinoma. Sydney: Cancer Council Australia; 2015. Available from: http://wiki.cancer.org.au/australia/Guidelines:Barrett%27s.
52. Whiteman DC, Appleyard M, Bahin FF, Bobryshev YV, Bourke MJ, Brown I, Chung A, Clouston A, Dickins E, Emery J, Eslick GD, Gordon LG, Grimpen F, Hebbard G, Holliday L, Hourigan L, Kendall BJ, Lee EY, Levert A, Lord RV, Lord SJ, Maule D, Moss A, Norton I, Olver I, Pavey D, Raftopoulos S, Rajendra S, Schoeman M, Singh R, Sitas F, Smithers BM, Taylor A, Thomas ML, Thomson I, To H, von Dincklage J, Vuletich C, Watson DI, Yusoff IF. Australian clinical practice guidelines for the diagnosis and management of Barrett's Esophagus and Early Esophageal Adenocarcinoma. Journal of gastroenterology and hepatology. 2015.
53. Schulmann K, Sterian A, Berki A, Yin J, Sato F, Xu Y, Olaru A, Wang S, Mori Y, Deacu E, Hamilton J, Kan T, Krasna MJ, Beer DG, Pepe MS, Abraham JM, Feng Z, Schmiegel W, Greenwald BD, Meltzer SJ. Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk. Oncogene. 2005;24(25):4138-48.
54. Wong DJ, Barrett MT, Stoger R, Emond MJ, Reid BJ. p16INK4a promoter is hypermethylated at a high frequency in esophageal adenocarcinomas. Cancer research. 1997;57(13):2619-22.
55. Wong DJ, Paulson TG, Prevo LJ, Galipeau PC, Longton G, Blount PL, Reid BJ. p16(INK4a) lesions are common, early abnormalities that undergo clonal expansion in Barrett's metaplastic epithelium. Cancer research. 2001;61(22):8284-9.
56. Eads CA, Lord RV, Kurumboor SK, Wickramasinghe K, Skinner ML, Long TI, Peters JH, DeMeester TR, Danenberg KD, Danenberg PV, Laird PW, Skinner KA. Fields of aberrant CpG island hypermethylation in Barrett's esophagus and associated adenocarcinoma. Cancer research. 2000;60(18):5021-6.
57. Corn PG, Heath EI, Heitmiller R, Fogt F, Forastiere AA, Herman JG, Wu TT. Frequent hypermethylation of the 5' CpG island of E-cadherin in esophageal adenocarcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2001;7(9):2765-9.
58. Bian YS, Osterheld MC, Fontolliet C, Bosman FT, Benhattar J. p16 inactivation by methylation of the CDKN2A promoter occurs early during neoplastic progression in Barrett's esophagus. Gastroenterology. 2002;122(4):1113-21.
59. Jin Z, Mori Y, Yang J, Sato F, Ito T, Cheng Y, Paun B, Hamilton JP, Kan T, Olaru A, David S, Agarwal R, Abraham JM, Beer D, Montgomery E, Meltzer SJ. Hypermethylation of the nel-like 1 gene is a common and early event and is associated with poor prognosis in early-stage esophageal adenocarcinoma. Oncogene. 2007;26(43):6332-40.
60. Jin Z, Olaru A, Yang J, Sato F, Cheng Y, Kan T, Mori Y, Mantzur C, Paun B, Hamilton JP, Ito T, Wang S, David S, Agarwal R, Beer DG, Abraham JM, Meltzer SJ. Hypermethylation of tachykinin-1 is a potential biomarker in

human esophageal cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13(21):6293-300.
61. Jin Z, Mori Y, Hamilton JP, Olaru A, Sato F, Yang J, Ito T, Kan T, Agarwal R, Meltzer SJ. Hypermethylation of the somatostatin promoter is a common, early event in human esophageal carcinogenesis. Cancer. 2008;112(1):43-9.
62. Jin Z, Hamilton JP, Yang J, Mori Y, Olaru A, Sato F, Ito T, Kan T, Cheng Y, Paun B, David S, Beer DG, Agarwal R, Abraham JM, Meltzer SJ. Hypermethylation of the AKAP12 promoter is a biomarker of Barrett's-associated esophageal neoplastic progression. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008;17(1):111-7.
63. Jin Z, Cheng Y, Olaru A, Kan T, Yang J, Paun B, Ito T, Hamilton JP, David S, Agarwal R, Selaru FM, Sato F, Abraham JM, Beer DG, Mori Y, Shimada Y, Meltzer SJ. Promoter hypermethylation of CDH13 is a common, early event in human esophageal adenocarcinogenesis and correlates with clinical risk factors. International journal of cancer Journal international du cancer. 2008;123(10):2331-6.
64. Alvi MA, Liu X, O'Donovan M, Newton R, Wernisch L, Shannon NB, Shariff K, di Pietro M, Bergman JJ, Ragunath K, Fitzgerald RC. DNA methylation as an adjunct to histopathology to detect prevalent, inconspicuous dysplasia and early-stage neoplasia in Barrett's esophagus. Clinical cancer research : an official journal of the American Association for Cancer Research. 2013;19(4):878-88.
65. Li X, Zhou F, Jiang C, Wang Y, Lu Y, Yang F, Wang N, Yang H, Zheng Y, Zhang J. Identification of a DNA methylome profile of esophageal squamous cell carcinoma and potential plasma epigenetic biomarkers for early diagnosis. PloS one. 2014;9(7):e103162.
66. Martin KJ, Fournier MV, Reddy GP, Pardee AB. A need for basic research on fluid-based early detection biomarkers. Cancer research. 2010;70(13):5203-6.
67. Longuespee R, Boyon C, Desmons A, Vinatier D, Leblanc E, Farre I, Wisztorski M, Ly K, D'Anjou F, Day R, Fournier I, Salzet M. Ovarian cancer molecular pathology. Cancer metastasis reviews. 2012;31(3-4):713-32.
68. Dorr M, Holzel D, Schubert-Fritschle G, Engel J, Schlesinger-Raab A. Changes in Prognostic and Therapeutic Parameters in Prostate Cancer from an Epidemiological View over 20 Years. Oncology research and treatment. 2015.
69. Vatandoost N, Ghanbari J, Mojaver M, Avan A, Ghayour-Mobarhan M, Nedaeinia R, Salehi R. Early detection of colorectal cancer: from conventional methods to novel biomarkers. Journal of cancer research and clinical oncology. 2015.
70. Qu Y, Dang S, Hou P. Gene methylation in gastric cancer. Clinica chimica acta; international journal of clinical chemistry. 2013;424:53-65.
71. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, Kaper F, Dawson S-J, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD, Rosenfeld N. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. Science translational medicine. 2012;4(136):136ra68-ra68.

72. Murtaza M, Dawson S-J, Pogrebniak K, Rueda OM, Provenzano E, Grant J, Chin S-F, Tsui DWY, Marass F, Gale D, Ali HR, Shah P, Contente-Cuomo T, Farahani H, Shumansky K, Kingsbury Z, Humphray S, Bentley D, Shah SP, Wallis M, Rosenfeld N, Caldas C. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. Nature communications. 2015;6.
73. Melnikov A, Scholtens D, Godwin A, Levenson V. Differential methylation profile of ovarian cancer in tissues and plasma. The Journal of molecular diagnostics : JMD. 2009;11(1):60-5.
74. Iyer P, Zekri AR, Hung CW, Schiefelbein E, Ismail K, Hablas A, Seifeldin IA, Soliman AS. Concordance of DNA methylation pattern in plasma and tumor DNA of Egyptian hepatocellular carcinoma patients. Experimental and molecular pathology. 2010;88(1):107-11.
75. Hoffmann AC, Vallbohmer D, Prenzel K, Metzger R, Heitmann M, Neiss S, Ling F, Holscher AH, Schneider PM, Brabender J. Methylated DAPK and APC promoter DNA detection in peripheral blood is significantly associated with apparent residual tumor and outcome. Journal of cancer research and clinical oncology. 2009;135(9):1231-7.
76. Kawakami K, Brabender J, Lord RV, Groshen S, Greenwald BD, Krasna MJ, Yin J, Fleisher AS, Abraham JM, Beer DG, Sidransky D, Huss HT, Demeester TR, Eads C, Laird PW, Ilson DH, Kelsen DP, Harpole D, Moore MB, Danenberg KD, Danenberg PV, Meltzer SJ. Hypermethylated APC DNA in plasma and prognosis of patients with esophageal adenocarcinoma. Journal of the National Cancer Institute. 2000;92(22):1805-11.
77. Peters JH, Clark GW, Ireland AP, Chandrasoma P, Smyrk TC, DeMeester TR. Outcome of adenocarcinoma arising in Barrett's esophagus in endoscopically surveyed and nonsurveyed patients. The Journal of thoracic and cardiovascular surgery. 1994;108(5):813-21; discussion 21-2.
78. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. Journal of the National Cancer Institute. 2001;93(14):1054-61.
79. Monis PT, Giglio S, Saint CP. Comparison of SYTO9 and SYBR Green I for real-time polymerase chain reaction and investigation of the effect of dye concentration on amplification and DNA melting curve analysis. Analytical biochemistry. 2005;340(1):24-34.
80. Odze RD. Diagnosis and grading of dysplasia in Barrett's oesophagus. Journal of Clinical Pathology. 2006;59(10):1029-38.
81. Fitzgerald RC, di Pietro M, Ragunath K, Ang Y, Kang JY, Watson P, Trudgill N, Patel P, Kaye PV, Sanders S, O'Donovan M, Bird-Lieberman E, Bhandari P, Jankowski JA, Attwood S, Parsons SL, Loft D, Lagergren J, Moayyedi P, Lyratzopoulos G, de Caestecker J. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. Gut. 2014;63(1):7-42.
82. Reid BJ, Haggitt RC, Rubin CE, Roth G, Surawicz CM, Van Belle G, Lewin K, Weinstein WM, Antonioli DA, Goldman H, et al. Observer variation in the diagnosis of dysplasia in Barrett's esophagus. Hum Pathol. 1988;19(2):166-78.

83. Montgomery E. Is there a way for pathologists to decrease interobserver variability in the diagnosis of dysplasia? Archives of pathology & laboratory medicine. 2005;129(2):174-6.
84. Montgomery E, Bronner MP, Goldblum JR, Greenson JK, Haber MM, Hart J, Lamps LW, Lauwers GY, Lazenby AJ, Lewin DN, Robert ME, Toledano AY, Shyr Y, Washington K. Reproducibility of the diagnosis of dysplasia in Barrett esophagus: a reaffirmation. Hum Pathol. 2001;32(4):368-78.
85. Nicholson AM, Graham TA, Simpson A, Humphries A, Burch N, Rodriguez-Justo M, Novelli M, Harrison R, Wright NA, McDonald SA, Jankowski JA. Barrett's metaplasia glands are clonal, contain multiple stem cells and share a common squamous progenitor. Gut. 2012;61(10):1380-9.
86. Xian W, Ho KY, Crum CP, McKeon F. Cellular origin of Barrett's esophagus: controversy and therapeutic implications. Gastroenterology. 2012;142(7):1424-30.
87. Plum PS, Bollschweiler E, Holscher AH, Warnecke-Eberz U. Novel diagnostic and prognostic biomarkers in esophageal cancer. Expert opinion on medical diagnostics. 2013;7(6):557-71.
88. Warton K, Lin V, Navin T, Armstrong NJ, Kaplan W, Ying K, Gloss B, Mangs H, Nair SS, Hacker NF, Sutherland RL, Clark SJ, Samimi G. Methylation-capture and Next-Generation Sequencing of free circulating DNA from human plasma. BMC genomics. 2014;15:476.
89. Tao YF, Hu SY, Lu J, Cao L, Zhao WL, Xiao PF, Xu LX, Li ZH, Wang NN, Du XJ, Sun LC, Zhao H, Fang F, Su GH, Li YH, Li YP, Xu YY, Ni J, Wang J, Feng X, Pan J. Zinc finger protein 382 is downregulated by promoter hypermethylation in pediatric acute myeloid leukemia patients. International journal of molecular medicine. 2014;34(6):1505-15.
90. Nones K, Waddell N, Wayte N, Patch AM, Bailey P, Newell F, Holmes O, Fink JL, Quinn MC, Tang YH, Lampe G, Quek K, Loffler KA, Manning S, Idrisoglu S, Miller D, Xu Q, Waddell N, Wilson PJ, Bruxner TJ, Christ AN, Harliwong I, Nourse C, Nourbakhsh E, Anderson M, Kazakoff S, Leonard C, Wood S, Simpson PT, Reid LE, Krause L, Hussey DJ, Watson DI, Lord RV, Nancarrow D, Phillips WA, Gotley D, Smithers BM, Whiteman DC, Hayward NK, Campbell PJ, Pearson JV, Grimmond SM, Barbour AP. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. Nature communications. 2014;5:5224.
91. Brock MV, Gou M, Akiyama Y, Muller A, Wu TT, Montgomery E, Deasel M, Germonpre P, Rubinson L, Heitmiller RF, Yang SC, Forastiere AA, Baylin SB, Herman JG. Prognostic importance of promoter hypermethylation of multiple genes in esophageal adenocarcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2003;9(8):2912-9.
92. Hamilton JP, Sato F, Greenwald BD, Suntharalingam M, Krasna MJ, Edelman MJ, Doyle A, Berki AT, Abraham JM, Mori Y, Kan T, Mantzur C, Paun B, Wang S, Ito T, Jin Z, Meltzer SJ. Promoter methylation and response to chemotherapy and radiation in esophageal cancer. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2006;4(6):701-8.
93. Kuester D, El-Rifai W, Peng D, Ruemmele P, Kroeckel I, Peters B, Moskaluk CA, Stolte M, Monkemuller K, Meyer F, Schulz HU, Hartmann A,

Roessner A, Schneider-Stock R. Silencing of MGMT expression by promoter hypermethylation in the metaplasia-dysplasia-carcinoma sequence of Barrett's esophagus. Cancer letters. 2009;275(1):117-26.
94. Moarii M, Boeva V, Vert J-P, Reyal F. Changes in correlation between promoter methylation and gene expression in cancer. BMC genomics. 2015;16(1):1-14.
95. Mobius C, Stein HJ, Becker I, Feith M, Theisen J, Gais P, Jutting U, Siewert JR. The 'angiogenic switch' in the progression from Barrett's metaplasia to esophageal adenocarcinoma. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2003;29(10):890-4.
96. Couvelard A, Paraf F, Gratio V, Scoazec JY, Henin D, Degott C, Flejou JF. Angiogenesis in the neoplastic sequence of Barrett's oesophagus. Correlation with VEGF expression. The Journal of pathology. 2000;192(1):14-8.
97. Lord RV, Park JM, Wickramasinghe K, DeMeester SR, Oberg S, Salonga D, Singer J, Peters JH, Danenberg KD, Demeester TR, Danenberg PV. Vascular endothelial growth factor and basic fibroblast growth factor expression in esophageal adenocarcinoma and Barrett esophagus. The Journal of thoracic and cardiovascular surgery. 2003;125(2):246-53.
98. Ahn HJ, Lee DS. Helicobacter pylori in gastric carcinogenesis. World journal of gastrointestinal oncology. 2015;7(12):455-65.
99. Souza RF, Krishnan K, Spechler SJ. Acid, bile, and CDX: the ABCs of making Barrett's metaplasia. American journal of physiology Gastrointestinal and liver physiology. 2008;295(2):G211-8.
100. Yamamoto Y, Wang X, Bertrand D, Kern F, Zhang T, Duleba M, Srivastava S, Khor CC, Hu Y, Wilson LH, Blaszyk H, Rolshud D, Teh M, Liu J, Howitt BE, Vincent M, Crum CP, Nagarajan N, Ho KY, McKeon F, Xian W. Mutational spectrum of Barrett/'s stem cells suggests paths to initiation of a precancerous lesion. Nature communications. 2016;7.
101. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, Weinberg RA. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nature genetics. 2008;40(5):499-507.
102. Melino G. p63 is a suppressor of tumorigenesis and metastasis interacting with mutant p53. Cell death and differentiation. 2011;18(9):1487-99.
103. Park BJ, Lee SJ, Kim JI, Lee SJ, Lee CH, Chang SG, Park JH, Chi SG. Frequent alteration of p63 expression in human primary bladder carcinomas. Cancer research. 2000;60(13):3370-4.
104. Wang X, Ouyang H, Yamamoto Y, Kumar PA, Wei TS, Dagher R, Vincent M, Lu X, Bellizzi AM, Ho KY, Crum CP, Xian W, McKeon F. Residual embryonic cells as precursors of a Barrett's-like metaplasia. Cell. 2011;145(7):1023-35.
105. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC bioinformatics. 2010;11:587.
106. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting

genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.
107. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1[alpha]-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics. 2003;34(3):267-73.
108. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44-57.
109. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research. 2009;37(1):1-13.
110. Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, Suchiman HE, Lameijer EW, Putter H, van den Akker EB, Zhang Y, Jukema JW, Slagboom PE, Meulenbelt I, Heijmans BT. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics & chromatin. 2013;6(1):26.
111. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011;3(6):771-84.
112. Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, Palotie A, Coffey AJ. A Comparison of the Whole Genome Approach of MeDIP-Seq to the Targeted Approach of the Infinium HumanMethylation450 BeadChip® for Methylome Profiling. PloS one. 2012;7(11):e50233.
113. Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, Lehmann U. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. BMC research notes. 2012;5:210.
114. An J, Gonzalez-Avalos E, Chawla A, Jeong M, Lopez-Moyado IF, Li W, Goodell MA, Chavez L, Ko M, Rao A. Acute loss of TET function results in aggressive myeloid cancer in mice. Nature communications. 2015;6:10071.
115. Mahauad-Fernandez WD, Borcherding NC, Zhang W, Okeoma CM. Bone Marrow Stromal Antigen 2 (BST-2) DNA Is Demethylated in Breast Tumors and Breast Cancer Cells. PloS one. 2015;10(4):e0123931.
116. Sayeed A, Luciani-Torres G, Meng Z, Bennington JL, Moore DH, Dairkee SH. Aberrant Regulation of the BST2 (Tetherin) Promoter Enhances Cell Proliferation and Apoptosis Evasion in High Grade Breast Cancer Cells. PloS one. 2013;8(6).
117. Jones PH, Mahauad-Fernandez WD, Madison MN, Okeoma CM. BST-2/tetherin is overexpressed in mammary gland and tumor tissues in MMTV-induced mammary cancer. Virology. 2013;444(0):124-39.
118. Cai D, Cao J, Li Z, Zheng X, Yao Y, Li W, Yuan Z. Up-regulation of bone marrow stromal protein 2 (BST2) in breast cancer with bone metastasis. BMC cancer. 2009;9(1):1-10.
119. Fang K-H, Kao H-K, Chi L-M, Liang Y, Liu S-C, Hseuh C, Liao C-T, Yen T-C, Yu J-S, Chang K-P. Overexpression of BST2 is associated with

nodal metastasis and poorer prognosis in oral cavity cancer. The Laryngoscope. 2014;124(9):E354-E60.
120. Wang W, Nishioka Y, Ozaki S, Jalili A, Abe S, Kakiuchi S, Kishuku M, Minakuchi K, Matsumoto T, Sone S. HM1.24 (CD317) is a novel target against lung cancer for immunotherapy using anti-HM1.24 antibody. Cancer immunology, immunotherapy : CII. 2009;58(6):967-76.
121. Tai YT, Horton HM, Kong SY, Pong E, Chen H, Cemerski S, Bernett MJ, Nguyen DH, Karki S, Chu SY, Lazar GA, Munshi NC, Desjarlais JR, Anderson KC, Muchhal US. Potent in vitro and in vivo activity of an Fc-engineered humanized anti-HM1.24 antibody against multiple myeloma via augmented effector function. Blood. 2012;119(9):2074-82.
122. Staudinger M, Glorius P, Burger R, Kellner C, Klausz K, Gunther A, Repp R, Klapper W, Gramatzki M, Peipp M. The novel immunotoxin HM1.24-ETA' induces apoptosis in multiple myeloma cells. Blood Cancer J. 2014;4:11.
123. Wainwright DA, Balyasnikova IV, Han Y, Lesniak MS. The expression of BST2 in human and experimental mouse brain tumors. Experimental and molecular pathology. 2011;91(1):440-6.
124. Milutin Gasperov N, Farkas SA, Nilsson TK, Grce M. Epigenetic activation of immune genes in cervical cancer. Immunology letters. 2014;162(2 Pt B):256-7.
125. Yokoyama T, Enomoto T, Serada S, Morimoto A, Matsuzaki S, Ueda Y, Yoshino K, Fujita M, Kyo S, Iwahori K, Fujimoto M, Kimura T, Naka T. Plasma membrane proteomics identifies bone marrow stromal antigen 2 as a potential therapeutic target in endometrial cancer. International journal of cancer Journal international du cancer. 2013;132(2):472-84.
126. Wong YF, Cheung TH, Lo KW, Yim SF, Siu NS, Chan SC, Ho TW, Wong KW, Yu MY, Wang VW, Li C, Gardner GJ, Bonome T, Johnson WB, Smith DI, Chung TK, Birrer MJ. Identification of molecular markers and signaling pathway in endometrial cancer in Hong Kong Chinese women by genome-wide gene expression profiling. Oncogene. 2007;26(13):1971-82.
127. Mahauad-Fernandez WD, Okeoma CM. The role of BST-2/Tetherin in host protection and disease manifestation. Immunity, Inflammation and Disease. 2016;4(1):4-23.
128. Brentani H, Caballero OL, Camargo AA, da Silva AM, da Silva WA, Neto ED, Grivet M, Gruber A, Guimaraes PEM, Hide W, Iseli C, Jongeneel CV, Kelso J, Nagai MA, Ojopi EPB, Osorio EC, Reis EMR, Riggins GJ, Simpson AJG, de Souza S, Stevenson BJ, Strausberg RL, Tajara EH, Verjovski-Almeida S, Consortium THCGPCGAPA, Consortium THCGPS. The generation and utilization of a cancer-oriented representation of the human transcriptome by using expressed sequence tags. Proceedings of the National Academy of Sciences. 2003;100(23):13418-23.
129. Segal E, Friedman N, Koller D, Regev A. A module map showing conditional activity of expression modules in cancer. Nature genetics. 2004;36(10):1090-8.
130. Barzi A, Lenz HJ. Angiogenesis-related agents in esophageal cancer. Expert opinion on biological therapy. 2012;12(10):1335-45.
131. Dreikhausen L, Blank S, Sisic L, Heger U, Weichert W, Jäger D, Bruckner T, Giese N, Grenacher L, Falk C, Ott K, Schmidt T. Association of

angiogenic factors with prognosis in esophageal cancer. BMC cancer. 2015;15(1):1-11.
132. Hedner C, Borg D, Nodin B, Karnevi E, Jirstrom K, Eberhard J. Expression and Prognostic Significance of Human Epidermal Growth Factor Receptors 1 and 3 in Gastric and Esophageal Adenocarcinoma. PloS one. 2016;11(2):e0148101.
133. Kestens C, Siersema PD, Offerhaus GJ, van Baal JW. BMP4 Signaling Is Able to Induce an Epithelial-Mesenchymal Transition-Like Phenotype in Barrett's Esophagus and Esophageal Adenocarcinoma through Induction of SNAIL2. PloS one. 2016;11(5):e0155754.
134. Gabay M, Li Y, Felsher DW. MYC activation is a hallmark of cancer initiation and maintenance. Cold Spring Harbor perspectives in medicine. 2014;4(6).
135. Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM, Harrison DJ, Meehan RR, Sims AH, Ramsahoye BH. Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(11):4364-9.
136. Sproul D, Meehan RR. Genomic insights into cancer-associated aberrant CpG island hypermethylation. Briefings in functional genomics. 2013;12(3):174-90.
137. Sproul D, Kitchen RR, Nestor CE, Dixon JM, Sims AH, Harrison DJ, Ramsahoye BH, Meehan RR. Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. Genome biology. 2012;13(10):R84.
138. Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, Pikarski E, Young RA, Niveleau A, Cedar H, Simon I. Evidence for an instructive mechanism of de novo methylation in cancer cells. Nature genetics. 2006;38(2):149-53.
139. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. Nature. 2005;435(7042):637-45.
140. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome biology. 2013;14(3):R21.
141. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature methods. 2012;9(3):215-6.
142. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Muller F, Forrest AR, Carninci P, Rehli M, Sandelin A. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507(7493):455-61.
143. Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M. DNA methylation: bisulphite modification and analysis. Nature protocols. 2006;1(5):2353-64.

144. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. Bioinformatics (Oxford, England). 2002;18(11):1427-31.
145. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research. 2007;35(Web Server issue):W71-4.
146. Breslauer KJ, Frank R, Blocker H, Marky LA. Predicting DNA duplex stability from the base sequence. Proceedings of the National Academy of Sciences of the United States of America. 1986;83(11):3746-50.
147. Kanemoto M, Shirahata M, Nakauma A, Nakanishi K, Taniguchi K, Kukita Y, Arakawa Y, Miyamoto S, Kato K. Prognostic prediction of glioblastoma by quantitative assessment of the methylation status of the entire MGMT promoter region. BMC cancer. 2014;14:641.
148. Chen C, Hua H, Han C, Cheng Y, Cheng Y, Wang Z, Bao J. Prognosis value of MGMT promoter methylation for patients with lung cancer: a meta-analysis. International journal of clinical and experimental pathology. 2015;8(9):11560-4.
149. Fornaro L, Vivaldi C, Caparello C, Musettini G, Baldini E, Masi G, Falcone A. Pharmacoepigenetics in gastrointestinal tumors: MGMT methylation and beyond. Frontiers in bioscience (Elite edition). 2016;8:170-80.
150. Sun Y, Li S, Shen K, Ye S, Cao D, Yang J. DAPK1, MGMT and RARB promoter methylation as biomarkers for high-grade cervical lesions. International journal of clinical and experimental pathology. 2015;8(11):14939-45.
151. Smith E, De Young NJ, Pavey SJ, Hayward NK, Nancarrow DJ, Whiteman DC, Smithers BM, Ruszkiewicz AR, Clouston AD, Gotley DC, Devitt PG, Jamieson GG, Drew PA. Similarity of aberrant DNA methylation in Barrett's esophagus and esophageal adenocarcinoma. Molecular cancer. 2008;7:75.
152. Jiang Z, Hu J, Li X, Jiang Y, Zhou W, Lu D. Expression analyses of 27 DNA repair genes in astrocytoma by TaqMan low-density array. Neuroscience letters. 2006;409(2):112-7.
153. Bernstein C, Bernstein H. Epigenetic reduction of DNA repair in progression to gastrointestinal cancer. World journal of gastrointestinal oncology. 2015;7(5):30-46.
154. Louro R, Nakaya HI, Paquola AC, Martins EA, da Silva AM, Verjovski-Almeida S, Reis EM. RASL11A, member of a novel small monomeric GTPase gene family, is down-regulated in prostate tumors. Biochemical and biophysical research communications. 2004;316(3):618-27.
155. Krauss G. Biochemistry of Signal Transduction and Regulation: Wiley; 2008.
156. Valle-Rios R, Maravillas-Montero Jé L, Burkhardt AM, Martinez C, Buhren BA, Homey B, Gerber PA, Robinson O, Hevezi P, Zlotnik A. Isthmin 1 Is a Secreted Protein Expressed in Skin, Mucosal Tissues, and NK, NKT, and Th17 Cells. Journal of Interferon & Cytokine Research. 2014;34(10):795-801.
157. Chen M, Zhang Y, Yu VC, Chong YS, Yoshioka T, Ge R. Isthmin targets cell-surface GRP78 and triggers apoptosis via induction of mitochondrial dysfunction. Cell death and differentiation. 2014;21(5):797-810.

158. Xiang W, Ke Z, Zhang Y, Cheng GH, Irwan ID, Sulochana KN, Potturi P, Wang Z, Yang H, Wang J, Zhuo L, Kini RM, Ge R. Isthmin is a novel secreted angiogenesis inhibitor that inhibits tumour growth in mice. Journal of cellular and molecular medicine. 2011;15(2):359-74.
159. Ch'ang HJ. Optimal combination of antiangiogenic therapy for hepatocellular carcinoma. World journal of hepatology. 2015;7(16):2029-40.
160. Longoria TC, Tewari KS. Pharmacologic management of advanced cervical cancer: antiangiogenesis therapy and immunotherapeutic considerations. Drugs. 2015;75(16):1853-65.
161. Shi L, Zhou J, Wu J, Shen Y, Li X. Anti-Angiogenic Therapy: Strategies to Develop Potent VEGFR-2 Tyrosine Kinase Inhibitors and Future Prospect. Current medicinal chemistry. 2016.
162. Yuan B, Xian R, Ma J, Chen Y, Lin C, Song Y. Isthmin inhibits glioma growth through antiangiogenesis in vivo. Journal of neuro-oncology. 2012;109(2):245-52.
163. Pongor L, Kormos M, Hatzis C, Pusztai L, Szabo A, Gyorffy B. A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. Genome medicine. 2015;7(1):104.
164. Freundt EC, Bidere N, Lenardo MJ. A different TIPE of immune homeostasis. Cell. 2008;133(3):401-2.
165. Sun H, Gong S, Carmody RJ, Hilliard A, Li L, Sun J, Kong L, Xu L, Hilliard B, Hu S, Shen H, Yang X, Chen YH. TIPE2, a negative regulator of innate and adaptive immunity that maintains immune homeostasis. Cell. 2008;133(3):415-26.
166. Moniz LS, Vanhaesebroeck B. A new TIPE of phosphoinositide regulator in cancer. Cancer cell. 2014;26(4):443-4.
167. Fayngerts SA, Wu J, Oxley CL, Liu X, Vourekas A, Cathopoulis T, Wang Z, Cui J, Liu S, Sun H, Lemmon MA, Zhang L, Shi Y, Chen YH. TIPE3 is the transfer protein of lipid second messengers that promote cancer. Cancer cell. 2014;26(4):465-78.
168. Luan YY, Yao YM, Sheng ZY. The tumor necrosis factor-alpha-induced protein 8 family in immune homeostasis and inflammatory cancer diseases. Journal of biological regulators and homeostatic agents. 2013;27(3):611-9.
169. Cui J, Hao C, Zhang W, Shao J, Zhang N, Zhang G, Liu S. Identical expression profiling of human and murine TIPE3 protein reveals links to its functions. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society. 2015;63(3):206-16.
170. Lou Y, Liu S. The TIPE (TNFAIP8) family in inflammation, immunity, and cancer. Molecular immunology. 2011;49(1-2):4-7.
171. Bieker JJ. Kruppel-like factors: three fingers in many pies. The Journal of biological chemistry. 2001;276(37):34355-8.
172. Kanazawa A, Kawamura Y, Sekine A, Iida A, Tsunoda T, Kashiwagi A, Tanaka Y, Babazono T, Matsuda M, Kawai K, Iiizumi T, Fujioka T, Imanishi M, Kaku K, Iwamoto Y, Kawamori R, Kikkawa R, Nakamura Y, Maeda S. Single nucleotide polymorphisms in the gene encoding Kruppel-like factor 7 are associated with type 2 diabetes. Diabetologia. 2005;48(7):1315-22.

173. Morris VA, Cummings CL, Korb B, Boaglio S, Oehler VG. Deregulated KLF4 Expression in Myeloid Leukemias Alters Cell Proliferation and Differentiation through MicroRNA and Gene Targets. Molecular and cellular biology. 2015;36(4):559-73.
174. Tien YT, Chang MH, Chu PY, Lin CS, Liu CH, Liao AT. Downregulation of the KLF4 transcription factor inhibits the proliferation and migration of canine mammary tumor cells. Veterinary journal (London, England : 1997). 2015;205(2):244-53.
175. He H, Li S, Hong Y, Zou H, Chen H, Ding F, Wan Y, Liu Z. Kruppel-like Factor 4 Promotes Esophageal Squamous Cell Carcinoma Differentiation by Up-regulating Keratin 13 Expression. The Journal of biological chemistry. 2015;290(21):13567-77.
176. Zhang D, Dai Y, Cai Y, Suo T, Liu H, Wang Y, Cheng Z, Liu H. KLF2 is downregulated in pancreatic ductal adenocarcinoma and inhibits the growth and migration of cancer cells. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2015.
177. Yan Q, Zhang W, Wu Y, Wu M, Zhang M, Shi X, Zhao J, Nan Q, Chen Y, Wang L, Cheng T, Li J, Bai Y, Liu S, Wang J. KLF8 promotes tumorigenesis, invasion and metastasis of colorectal cancer cells by transcriptional activation of FHL2. Oncotarget. 2015;6(28):25402-17.
178. Zhou S, Tang X, Tang F. Kruppel-like factor 17, a novel tumor suppressor: its low expression is involved in cancer metastasis. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2015.
179. Li S, Qin X, Cui A, Wu W, Ren L, Wang X. Low expression of KLF17 is associated with tumor invasion in esophageal carcinoma. International journal of clinical and experimental pathology. 2015;8(9):11157-63.
180. Sachdeva M, Dodd RD, Huang Z, Grenier C, Ma Y, Lev DC, Cardona DM, Murphy SK, Kirsch DG. Epigenetic silencing of Kruppel like factor-3 increases expression of pro-metastatic miR-182. Cancer letters. 2015;369(1):202-11.
181. Satoh K, Hata M, Yokota H. A Novel Member of the Leucine-Rich Repeat Superfamily Induced in Rat Astrocytes by β-Amyloid. Biochemical and biophysical research communications. 2002;290(2):756-62.
182. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, Wernerus H, Bjorling L, Ponten F. Towards a knowledge-based Human Protein Atlas. Nature biotechnology. 2010;28(12):1248-50.
183. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F. Proteomics. Tissue-based map of the human proteome. Science (New York, NY). 2015;347(6220):1260419.
184. Effendi K, Yamazaki K, Fukuma M, Sakamoto M. Overexpression of Leucine-Rich Repeat-Containing G Protein-Coupled Receptor 5 (LGR5) Represents a Typical Wnt/beta-Catenin Pathway-Activated Hepatocellular Carcinoma. Liver cancer. 2014;3(3-4):451-7.

185. Huang T, Qiu X, Xiao J, Wang Q, Wang Y, Zhang Y, Bai D. The prognostic role of Leucine-rich repeat-containing G-protein-coupled receptor 5 in gastric cancer: A systematic review with meta-analysis. Clinics and research in hepatology and gastroenterology. 2015.
186. Jang BG, Lee BL, Kim WH. Prognostic significance of leucine-rich-repeat-containing G-protein-coupled receptor 5, an intestinal stem cell marker, in gastric carcinomas. Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association. 2015.
187. Satoh K, Hata M, Yokota H. High lib mRNA expression in breast carcinomas. DNA research : an international journal for rapid publication of reports on genes and genomes. 2004;11(3):199-203.
188. Schuetz CS, Bonin M, Clare SE, Nieselt K, Sotlar K, Walter M, Fehm T, Solomayer E, Riess O, Wallwiener D, Kurek R, Neubauer HJ. Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. Cancer research. 2006;66(10):5278-86.
189. Reynolds PA, Smolen GA, Palmer RE, Sgroi D, Yajnik V, Gerald WL, Haber DA. Identification of a DNA-binding site and transcriptional target for the EWS-WT1(+KTS) oncoprotein. Genes & development. 2003;17(17):2094-107.
190. Stanbrough M, Bubley GJ, Ross K, Golub TR, Rubin MA, Penning TM, Febbo PG, Balk SP. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. Cancer research. 2006;66(5):2815-25.
191. O'Prey J, Wilkinson S, Ryan KM. Tumor Antigen LRRC15 Impedes Adenoviral Infection: Implications for Virus-Based Cancer Therapy. Journal of Virology. 2008;82(12):5933-9.
192. Pallai R, Bhaskar A, Barnett-Bernodat N, Gallo-Ebert C, Pusey M, Nickels JT, Jr., Rice LM. Leucine-rich repeat-containing protein 59 mediates nuclear import of cancerous inhibitor of PP2A in prostate cancer cells. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2015;36(8):6383-90.
193. Chen B, Mao HH, Chen L, Zhang FL, Li K, Xue ZF. Loop-tail phenotype in heterozygous mice and neural tube defects in homozygous mice result from a nonsense mutation in the Vangl2 gene. Genetics and molecular research : GMR. 2013;12(3):3157-65.
194. De Marco P, Merello E, Piatelli G, Cama A, Kibar Z, Capra V. Planar cell polarity gene mutations contribute to the etiology of human neural tube defects in our population. Birth defects research Part A, Clinical and molecular teratology. 2014;100(8):633-41.
195. Murdoch JN, Damrau C, Paudyal A, Bogani D, Wells S, Greene ND, Stanier P, Copp AJ. Genetic interactions between planar cell polarity genes cause diverse neural tube defects in mice. Disease models & mechanisms. 2014;7(10):1153-63.
196. Hatakeyama J, Wald JH, Printsev I, Ho HY, Carraway KL, 3rd. Vangl1 and Vangl2: planar cell polarity components with a developing role in cancer. Endocrine-related cancer. 2014;21(5):R345-56.

197. Valastyan S, Weinberg Robert A. Tumor Metastasis: Molecular Insights and Evolving Paradigms. Cell. 2011;147(2):275-92.
198. Tano K, Shiota S, Collier J, Foote RS, Mitra S. Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O6-alkylguanine. Proceedings of the National Academy of Sciences of the United States of America. 1990;87(2):686-90.
199. Halford S, Rowan A, Sawyer E, Talbot I, Tomlinson I. O(6)-methylguanine methyltransferase in colorectal cancers: detection of mutations, loss of expression, and weak association with G:C>A:T transitions. Gut. 2005;54(6):797-802.
200. Hasina R, Surati M, Kawada I, Arif Q, Carey GB, Kanteti R, Husain AN, Ferguson MK, Vokes EE, Villaflor VM, Salgia R. O-6-methylguanine-deoxyribonucleic acid methyltransferase methylation enhances response to temozolomide treatment in esophageal cancer. Journal of Carcinogenesis. 2013;12.
201. Sanson M, Marie Y, Paris S, Idbaih A, Laffaire J, Ducray F, El Hallani S, Boisselier B, Mokhtari K, Hoang-Xuan K, Delattre JY. Isocitrate dehydrogenase 1 codon 132 mutation is an important prognostic biomarker in gliomas. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(25):4150-4.
202. Shah AK, Saunders NA, Barbour AP, Hill MM. Early diagnostic biomarkers for esophageal adenocarcinoma--the current state of play. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2013;22(7):1185-209.
203. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics. 2013;14(2):178-92.
204. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigo R, Hubbard TJ. GENCODE: the reference human genome annotation for The ENCODE Project. Genome research. 2012;22(9):1760-74.
205. Hulsen T, Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC genomics. 2008;9(1):1-6.
206. Weaver JM, Ross-Innes CS, Shannon N, Lynch AG, Forshew T, Barbera M, Murtaza M, Ong CA, Lao-Sirieix P, Dunning MJ, Smith L, Smith ML, Anderson CL, Carvalho B, O'Donovan M, Underwood TJ, May AP, Grehan N, Hardwick R, Davies J, Oloumi A, Aparicio S, Caldas C, Eldridge MD, Edwards PA, Rosenfeld N, Tavare S, Fitzgerald RC, Consortium O, Consortium O. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. Nature genetics. 2014;46(8):837-43.
207. Reid BJ, Paulson TG, Li X. Genetic Insights in Barrett's Esophagus and Esophageal Adenocarcinoma. Gastroenterology. 2015;149(5):1142-52.e3.

208. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, Bandla S, Imamura Y, Schumacher SE, Shefler E, McKenna A, Carter SL, Cibulskis K, Sivachenko A, Saksena G, Voet D, Ramos AH, Auclair D, Thompson K, Sougnez C, Onofrio RC, Guiducci C, Beroukhim R, Zhou Z, Lin L, Lin J, Reddy R, Chang A, Landrenau R, Pennathur A, Ogino S, Luketich JD, Golub TR, Gabriel SB, Lander ES, Beer DG, Godfrey TE, Getz G, Bass AJ. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nature genetics. 2013;45(5):478-86.
209. Xu H, DiCarlo J, Satya RV, Peng Q, Wang Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. BMC genomics. 2014;15(1):1-10.
210. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature biotechnology. 2013;31.
211. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics (Oxford, England). 2012;28:1811-7.
212. Hughes LA, Melotte V, de Schrijver J, de Maat M, Smit VT, Bovee JV, French PJ, van den Brandt PA, Schouten LJ, de Meyer T, van Criekinge W, Ahuja N, Herman JG, Weijenberg MP, van Engeland M. The CpG island methylator phenotype: what's in a name? Cancer research. 2013;73(19):5858-68.
213. Issa J-P. CpG island methylator phenotype in cancer. Nature reviews Cancer. 2004;4(12):988-93.
214. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, Campos C, Fabius AWM, Lu C, Ward PS, Thompson CB, Kaufman A, Guryanova O, Levine R, Heguy A, Viale A, Morris LGT, Huse JT, Mellinghoff IK, Chan TA. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature.483(7390):479-83.
215. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, Tallman MS, Sun Z, Wolniak K, Peeters JK, Liu W, Choe SE, Fantin VR, Paietta E, Lowenberg B, Licht JD, Godley LA, Delwel R, Valk PJ, Thompson CB, Levine RL, Melnick A. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer cell. 2010;18(6):553-67.
216. Shibata T, Kokubu A, Miyamoto M, Sasajima Y, Yamazaki N. Mutant IDH1 confers an in vivo growth in a melanoma cell line with BRAF mutation. The American journal of pathology. 2011;178(3):1395-402.
217. Reitman ZJ, Yan H. Isocitrate dehydrogenase 1 and 2 mutations in cancer: alterations at a crossroads of cellular metabolism. Journal of the National Cancer Institute. 2010;102(13):932-41.
218. Bleeker FE, Lamba S, Leenstra S, Troost D, Hulsebos T, Vandertop WP, Frattini M, Molinari F, Knowles M, Cerrato A, Rodolfo M, Scarpa A, Felicioni L, Buttitta F, Malatesta S, Marchetti A, Bardelli A. IDH1 mutations at residue p.R132 (IDH1(R132)) occur frequently in high-grade gliomas but not in other solid tumors. Human mutation. 2009;30(1):7-11.

219. An C, Choi IS, Yao JC, Worah S, Xie K, Mansfield PF, Ajani JA, Rashid A, Hamilton SR, Wu TT. Prognostic significance of CpG island methylator phenotype and microsatellite instability in gastric carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2005;11(2 Pt 1):656-63.
220. Maruyama R, Toyooka S, Toyooka KO, Harada K, Virmani AK, Zochbauer-Muller S, Farinas AJ, Vakar-Lopez F, Minna JD, Sagalowsky A, Czerniak B, Gazdar AF. Aberrant promoter methylation profile of bladder cancer and its relationship to clinicopathological features. Cancer research. 2001;61(24):8659-63.
221. Fang F, Turcan S, Rimner A, Kaufman A, Giri D, Morris LG, Shen R, Seshan V, Mo Q, Heguy A, Baylin SB, Ahuja N, Viale A, Massague J, Norton L, Vahdat LT, Moynahan ME, Chan TA. Breast cancer methylomes establish an epigenomic foundation for metastasis. Science translational medicine. 2011;3(75):75ra25.
222. Whitehall VL, Dumenil TD, McKeone DM, Bond CE, Bettington ML, Buttenshaw RL, Bowdler L, Montgomery GW, Wockner LF, Leggett BA. Isocitrate dehydrogenase 1 R132C mutation occurs exclusively in microsatellite stable colorectal cancers with the CpG island methylator phenotype. Epigenetics : official journal of the DNA Methylation Society. 2014;9(11):1454-60.
223. Liu Z, Zhao J, Chen XF, Li W, Liu R, Lei Z, Liu X, Peng X, Xu K, Chen J, Liu H, Zhou QH, Zhang HT. CpG island methylator phenotype involving tumor suppressor genes located on chromosome 3p in non-small cell lung cancer. Lung cancer (Amsterdam, Netherlands). 2008;62(1):15-22.
224. Strathdee G, Appleton K, Illand M, Millan DW, Sargent J, Paul J, Brown R. Primary ovarian carcinomas display multiple methylator phenotypes involving known tumor suppressor genes. The American journal of pathology. 2001;158(3):1121-7.
225. Ueki T, Toyota M, Sohn T, Yeo CJ, Issa JP, Hruban RH, Goggins M. Hypermethylation of multiple genes in pancreatic adenocarcinoma. Cancer research. 2000;60(7):1835-9.
226. Maruyama R, Toyooka S, Toyooka KO, Virmani AK, Zochbauer-Muller S, Farinas AJ, Minna JD, McConnell J, Frenkel EP, Gazdar AF. Aberrant promoter methylation profile of prostate cancers and its relationship to clinicopathological features. Clinical cancer research : an official journal of the American Association for Cancer Research. 2002;8(2):514-9.
227. Ogino S, Kawasaki T, Kirkner GJ, Loda M, Fuchs CS. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. The Journal of molecular diagnostics : JMD. 2006;8(5):582-8.
228. Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K, Hamilton SR, Issa JP. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(47):18654-9.
229. Curtin K, Slattery ML, Samowitz WS. CpG island methylation in colorectal cancer: past, present and future. Pathology research international. 2011;2011:902674.

230. Hinoue T, Weisenberger DJ, Pan F, Campan M, Kim M, Young J, Whitehall VL, Leggett BA, Laird PW. Analysis of the association between CIMP and BRAF in colorectal cancer by DNA methylation profiling. PloS one. 2009;4(12):e8357.
231. Selaru FM, David S, Meltzer SJ, Hamilton JP. Epigenetic Events in Gastrointestinal Cancer. Am J Gastroenterol. 2009;104(8):1910-2.
232. Li B, Liu W, Wang L, Li M, Wang J, Huang L, Huang P, Yuan Y. CpG island methylator phenotype associated with tumor recurrence in tumor-node-metastasis stage I hepatocellular carcinoma. Annals of surgical oncology. 2010;17(7):1917-26.
233. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2010;26(5):589-95.
234. Mallawaarachchi AC, Hort Y, Cowley MJ, McCabe MJ, Minoche A, Dinger ME, Shine J, Furlong TJ. Whole-genome sequencing overcomes pseudogene homology to diagnose autosomal dominant polycystic kidney disease. European journal of human genetics : EJHG. 2016;24(11):1584-90.
235. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43.
236. De Sousa SM, McCabe MJ, Wu K, Roscioli T, Gayevskiy V, Brook K, Rawlings L, Scott HS, Thompson TJ, Earls P, Gill AJ, Cowley MJ, Dinger ME, McCormack AI. Germline variants in familial pituitary tumour syndrome genes are common in young patients and families with additional endocrine tumours. European journal of endocrinology. 2017;176(5):635-44.
237. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics. 2014;46(3):310-5.
238. Kaz AM, Wong CJ, Luo Y, Virgin JB, Washington MK, Willis JE, Leidner RS, Chak A, Grady WM. DNA methylation profiling in Barrett's esophagus and esophageal adenocarcinoma reveals unique methylation signatures and molecular subclasses. Epigenetics : official journal of the DNA Methylation Society. 2011;6(12):1403-12.
239. Martins R, Bugalho MJ. Paragangliomas/Pheochromocytomas: clinically oriented genetic testing. International journal of endocrinology. 2014;2014:794187.
240. Rivenbark AG, Stolzenburg S, Beltran AS, Yuan X, Rots MG, Strahl BD, Blancafort P. Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics : official journal of the DNA Methylation Society. 2012;7(4):350-60.
241. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM, Antonarakis ES, Azad NS, Bardelli A, Brem H, Cameron JL, Lee CC, Fecher LA, Gallia GL, Gibbs P, Le D, Giuntoli RL, Goggins M, Hogarty MD, Holdhoff M, Hong SM, Jiao Y, Juhl HH, Kim JJ, Siravegna G, Laheru DA, Lauricella C, Lim M, Lipson EJ, Marie SK, Netto GJ, Oliner KS, Olivi A, Olsson L, Riggins GJ, Sartore-Bianchi A, Schmidt K, Shih I M, Oba-Shinjo SM, Siena S, Theodorescu D, Tie J, Harkins TT, Veronese S, Wang TL, Weingart JD, Wolfgang CL, Wood LD, Xing D, Hruban RH, Wu J, Allen PJ, Schmidt CM, Choti MA, Velculescu VE, Kinzler

KW, Vogelstein B, Papadopoulos N, Diaz LA, Jr. Detection of circulating tumor DNA in early- and late-stage human malignancies. Science translational medicine. 2014;6(224):224ra24.
242. Shaw JA, Stebbing J. Circulating free DNA in the management of breast cancer. Annals of translational medicine. 2014;2(1):3.

APPENDICES

Appendix 1: Website links to software, genome browsers and online tools used in this study

Software / Genome Browser / Online Tool and website link
Affymetrix® Expression Console™
http://www.affymetrix.com/support/learning/training_tutorials/tac_ec/index.affx
Bioconductor minfi package 3.2
http://bioconductor.org/packages/release/bioc/html/minfi.html
Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7
https://david.ncifcrf.gov
GenomeStudio v2011.1 with Methylation module 1.9.0
http://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html
HaplotypeCaller (Genome Analysis Toolkit (GATK) variant discovery tools)
https://www.broadinstitute.org/gatk/guide/tooldocs/org_broadinstitute_gatk_tools_walkers_haplotypecaller_HaplotypeCaller.php
Illumina Experiment Manager (IEM) v1.9
https://support.illumina.com/downloads/illumina-experiment-manager.html
Integrative Genomics Viewer (IGV)
https://www.broadinstitute.org/igv/
KAPA Library Quantification Data Analysis Template
https://www.kapabiosystems.com/document/kapa-library-quantification-data-analysis-template/
Linear Models for Microarray Analysis (LIMMA) package
https://bioconductor.org/packages/release/bioc/html/limma.html
MethPrimer: designing primer pairs for methylation PCRs
http://www.urogene.org/methprimer/
Multiple Primer Analyzer
https://www.thermofisher.com/au/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html
Transcriptome Analysis Console (TAC) software
http://www.affymetrix.com/support/learning/training_tutorials/tac_ec/index.affx
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) assembly
https://genome.ucsc.edu/cgi-bin/hgGateway

Appendix 2: Ethics documentation

Appendix 2.1 Human Research Ethics Committee confirmation letter HREC/13/SVH/344

St Vincent's Hospital

A facility of St Vincent's & Mater Health Sydney

St Vincent's Hospital Sydney Ltd
ABN 77 054 038 872
390 Victoria Street
Darlinghurst NSW 2010
Australia

T + 61 2 8382 1111
F + 61 2 9332 4142
www.stvincents.com.au

21 January 2014

A/Prof Reginald Lord
Suite 606
Level 6
St Vincent's Clinic
438 Victoria St
Darlinghurst NSW 2010

Dear Reginald

SVH File Number: 13/241
Project Title: Progression of Barrett's Oesophagus to Cancer Network NSW
HREC Reference Number: HREC/13/SVH/344

Thank you for your letter, dated **29 November 2013**, received **13 December 2013**, responding to issues raised regarding the above project, which was first considered by the St Vincent's Hospital HREC at its meeting held on **14 November 2013.** This HREC has been accredited by NSW Ministry of Health as a Lead HREC under the model for single ethical and scientific review and Certified by the NHMRC under the National model for Harmonisation of Multicentre Ethical Review (HoMER). This lead HREC is constituted and operates in accordance with the National Health and Medical Research Council's *National Statement on Ethical Conduct in Human Research* and the *CPMP/ICH Note for Guidance on Good Clinical Practice*. No HREC members with a conflict of interest were present for review of this project.

I am pleased to advise that the Committee at an Executive meeting on **14 January 2014** has granted ethical and scientific approval of the above **multi centre** project.

You are reminded that this letter constitutes *ETHICAL* and *SCIENTIFIC* approval only. You must not commence this research project at a site until a completed Site Specific Assessment Form/Access Request and associated documentation have been submitted to the site Research Governance Officer and Authorised. A copy of this letter must be forwarded to all site investigators for submission to the relevant Research Governance Officer.

The project is approved to be conducted at:

- St Vincent's Hospital
- Reginald V.N. Lord Private Rooms – St Vincent's Clinic – Suite 606, Level 6
- Westmead Hospital
- Nepean Hospital

If a new site(s) is to be added please inform the HREC in writing and submit a Site Specific Assessment Form (SSA) to the Research Governance Officer at the new site.

The following documentation has been reviewed and approved by the HREC:

- Participant Information Sheet and Consent Form (Information Booklet) version 1.6 dated 1 October 2013
- Participant Information Sheet and Consent Form - Consent For Research (Participant copy) version 2 dated 14 July 2008

- Participant Information Sheet and Consent Form – Consent For Research (Study copy) version 2 dated 14 July 2008
- Health Questionnaire version 2.2 dated 25 August 2008

The National Ethics Application Form (NEAF) document reviewed by the HREC was NEAF **AU/1/ECF418**.

Please note the following conditions of approval:

- HREC approval is valid for **5 years** from the date of the HREC Executive Committee meeting and expires on **14 January 2019.** The Co-ordinating Investigator is required to notify the HREC 6 months prior to this date if the project is expected to extend beyond the original approval date at which time the HREC will advise of the requirements for ongoing approval of the study.
- The Co-ordinating Investigator will provide an annual progress report beginning in **January 2015**, to the HREC as well as a final study report at the completion of the project in the specified format.
- The Co-ordinating Investigator will immediately report anything which might warrant review of ethical approval of the project in the specified format, including unforeseen events that might affect continued ethical acceptability of the project and any complaints made by study participants regarding the conduct of the study.
- Proposed changes to the research protocol, conduct of the research, or length of HREC approval will be provided to the HREC for review, in the specified format.
- The HREC will be notified, giving reasons, if the project is discontinued before the expected date of completion.
- Investigators holding an academic appointment (including conjoint appointments) and students undertaking a project as part of a University course may also be required to notify the relevant University HREC of the project. Investigators and students are advised to contact the relevant HREC to seek advice regarding their requirements.

Please note it is the responsibility of the sponsor or the co-ordinating investigator of the project to register this study on a publicly available online registry (eg. Australian Clinical Trial Registry www.actr.org.au).

Should you have any queries about your project please contact the Research Office, Tel: 8382-2075, email research@stvincents.com.au. The HREC Terms of Reference, Standard Operating Procedures, *National Statement on Ethical Conduct in Human Research* (2007) and the *CPMP/ICH Note for Guidance on Good Clinical Practice* and standard forms are available on the Research Office website: www.stvincents.com.au/researchoffice or internal: http://exwwwsvh.stvincents.com.au/researchoffice

Please quote **SVH File Number: 13/241** in all correspondence.

The HREC wishes you every success in your research.

Yours sincerely

Sarah Charlton
HREC Executive Officer
St Vincent's Research Office
Level 6 deLacy Building

cc. Angelique Levert
TRIM ref: D/2014/3469

Appendix 2.2 Site Specific Approval confirmation letter SSA/14/SVH/344

St Vincent's Hospital

A facility of St Vincent's
& Mater Health Sydney

St Vincent's Hospital Sydney Ltd
ABN 77 054 038 872
390 Victoria Street
Darlinghurst NSW 2010
Australia

T + 61 2 8382 1111
F + 61 2 9332 4142
www.stvincents.com.au

25 February 2014

A/Prof. Reginald Lord
Suite 606
Level 6, St Vincent's Clinic
438 Victoria Street
Darlinghurst NSW 2010

Dear Reginald

SVH File Number: 13/241
Project Title: Progression of Barrett's Oesophagus to Cancer Network NSW
HREC reference: HREC/13/SVH/344
SSA reference: SSA/14/SVH/46

Thank you for submitting an application for authorisation of this project. I am pleased to advise that the Director of Research, on **19 February 2014**, has granted authorisation for the above project to commence at **St Vincent's Hospital**.

Documents to be used at this site are:

- Participant Information Sheet and Consent Form (Information Booklet) version 1.6 dated 1 October 2013
- Participant Information Sheet and Consent Form - Consent For Research (Participant copy) version 2 dated 14 July 2008
- Participant Information Sheet and Consent Form – Consent For Research (Study copy) version 2 dated 14 July 2008
- Health Questionnaire version 2.2 dated 25 August 2008

The SSA form reviewed was: **AU/2/8786113.**

Please Note: Site authorisation will cease on the date of HREC expiry (14 January 2019).

The following conditions apply to this research project. These are additional to those conditions imposed by the Human Research Ethics Committee that granted ethical approval:

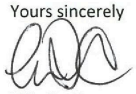
1. Please provide an amended Insurance Certificate with cover of $20 Million on any one claim (as per NSW Health Policy Directive PD2011_006) within 3 months of study commencement.
2. An annual progress report will be provided to the Research Governance Officer acknowledged by the LEAD HREC beginning in **January 2015.**
3. Proposed amendments to the research protocol or conduct of the research which may affect the ethical acceptability of the project, and are submitted to the lead HREC for review, are copied to the Research Governance Officer prior to implementation of the amendment on site.
4. Proposed amendments to the research protocol or conduct of the research which may affect the ongoing site acceptability of the project are to be submitted to the Research Governance Officer.

5. The relevant University HREC may require notification for projects that are undertaken by investigators holding an academic appointment (including conjoint appointments) or by students as part of a University course. This is the responsibility of the investigators.

Should you have any queries about your project please contact the Research Office, Ph 8382 2075, email research@stvincents.com.au. The HREC Terms of Reference, Standard Operating Procedures, *National Statement on Ethical Conduct in Human Research* (2007) and the *CPMP/ICH Note for Guidance on Good Clinical Practice* and standard forms are available on the Research Office website: www.stvincents.com.au/researchoffice or internal at http://exwwwsvh.stvincents.com.au/researchoffice.

Please quote **SVH file reference 13/241** and **HREC reference HREC/13/SVH/344** in any correspondence.

Yours sincerely



Ellie Pratt
Ethics and Research Governance Officer
St Vincent's Hospital Research Office
Level 6 deLacy

cc: Angelique Levert-Mignon
TRIM ref: D/2014/10142

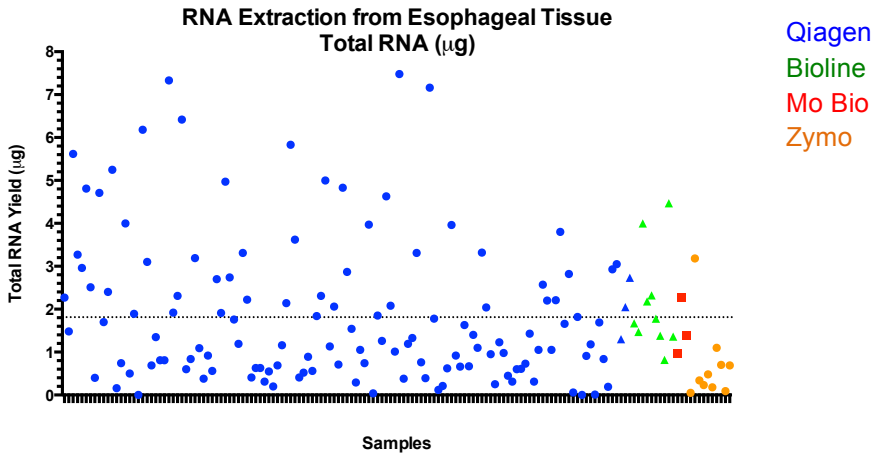
Appendix 3: Yield, purity, correlation and bias in RNA isolation from esophageal tissue

Yield, purity, correlation and biases in RNA isolation from esophageal tissue were examined following extraction from 154 tissue samples. Data includes samples extracted using trial kits (Bioline's Isolate II RNA Mini Kit, with extra column desalt, Mo Bio's UltraClean® Tissue & Cells RNA Isolation Kit and Zymo's Quick-RNA™ MiniPrep kit) as well as the selected Qiagen RNeasy Mini Kit (including protocol variations); chosen on the basis of superior absorbance ratios and RNA integrity number (RIN).

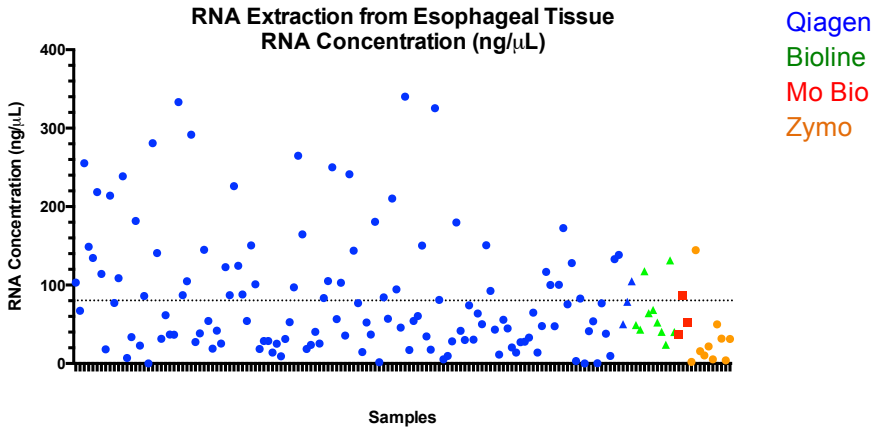
Appendix 3.1 RNA yields

(A) An average total RNA yield of 1.813 ± 1.632µg was obtained (median 1.315µg, maximum 7.480µg). (B) Average concentration of 80.36 ± 74.12ng/µL (median 54.05ng/µL, maximum 340.10ng/µL).

(A)



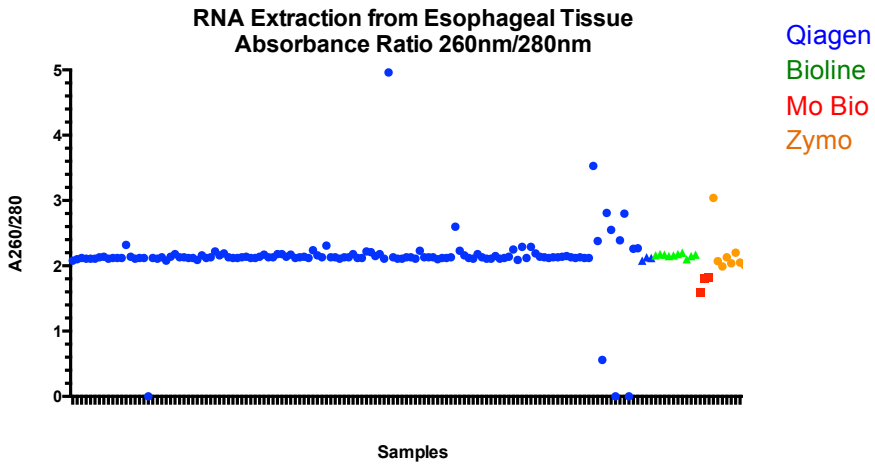
(B)



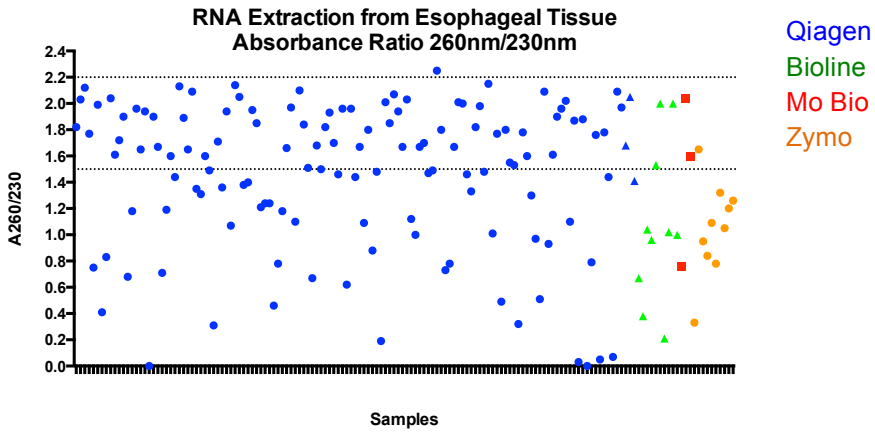
Appendix 3.2 Absorbance ratios

(A) Absorbance ratio A260/280 provides an indication of protein contamination. Pure RNA will have A260/280 ~2.0. An average A260/280 2.135 ± 0.4400 (median 2.130) was observed. Mo Bio's UltraClean® Tissue & Cells RNA Isolation Kit did not perform as well as the other kits in this regard. The cause for increased variation in the small group of Qiagen RNeasy Mini Kit samples (processed together in a single batch) was unable to be determined. (B) Average A260/230 of 1.416 ± 0.5626 (median 1.540) indicates some contamination, hypothesized due to high salt concentration from RNA*later*. Acceptable A260/230 range for downstream expression and methylation profiling was assigned 1.50 - 2.20.

(A)



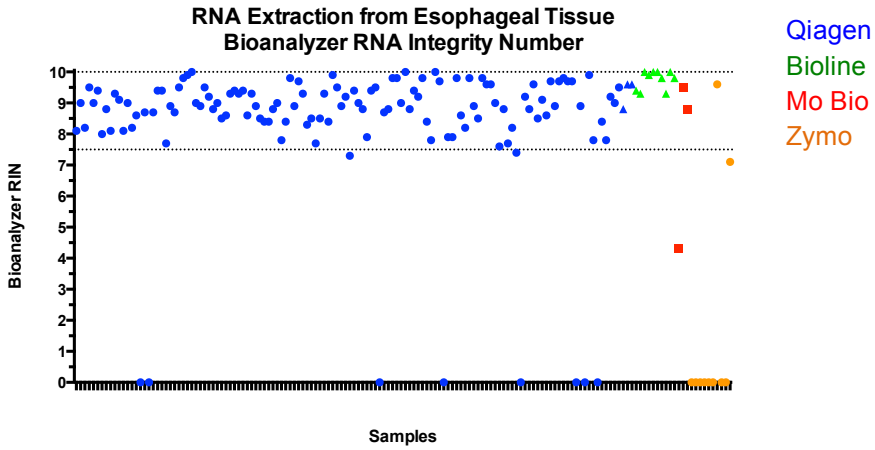
(B)



Appendix 3.3 Bioanalyzer RNA Integrity Number (RIN)

Bioanalyzer RIN (0 – 10) is used to estimate the integrity of total RNA using electrophoretic trace and takes into account the presence or absence of degradation products. We applied a cut-off of ≥7.5 for inclusion of the sample

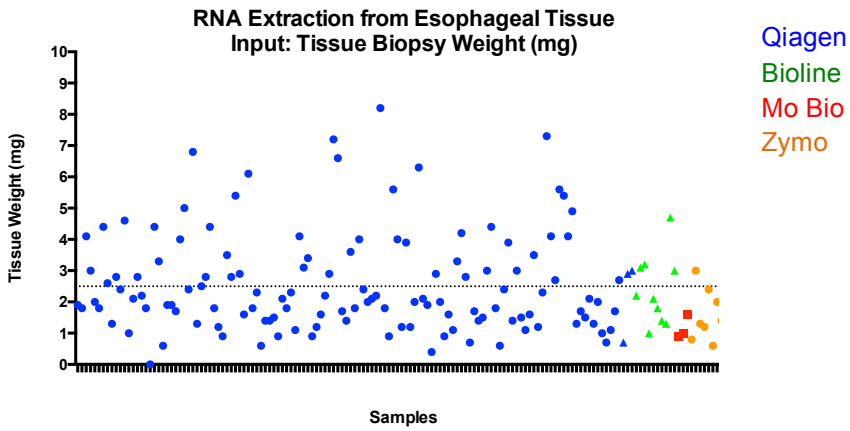
for genome-wide expression profiling. An average RIN of 8.029 ± 2.843 was obtained (median 8.90, maximum 10.00). Zymo's Quick-RNA™ MiniPrep kit was dismissed on this basis.



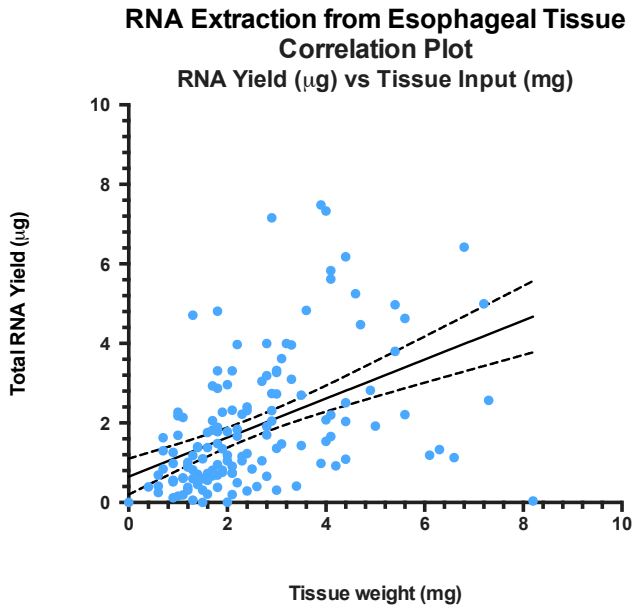
Appendix 3.4 Yield correlation with tissue input (input-output correlation)

(A) Average tissue biopsy input weight was 2.499 ± 1.560mg (median 2.050mg, maximum 8.200mg). (B) Correlation between tissue input and total RNA yield (input-output correlation). Linear regression line of best fit is plotted with 95% confidence band. Person correlation coefficient r^2 = 0.2189 (assumes Gaussian distribution).

(A)



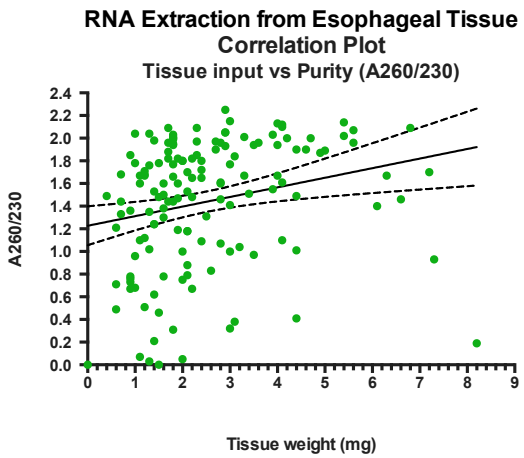
(B)



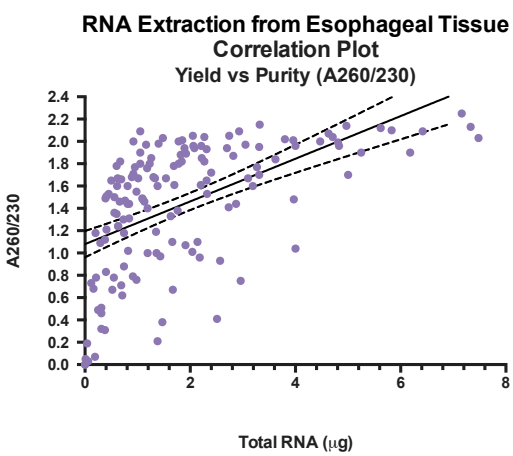
Appendix 3.5 Purity correlation (tissue input and yield)

(A) Correlation between purity (A260/230) and tissue input: investigating if a smaller initial biopsy results in poorer quality RNA. Linear regression line of best fit is plotted with 95% confidence band. Person correlation coefficient r^2 = 0.2346 (assumes Gaussian distribution). (B) Correlation between purity (A260/230) and total RNA yield (µg): investigating if a poor yield is synonymous with poor quality RNA. Linear regression line of best fit is plotted with 95% confidence band. Person correlation coefficient r^2 = 0.3085 (assumes Gaussian distribution).

(A)



(B)



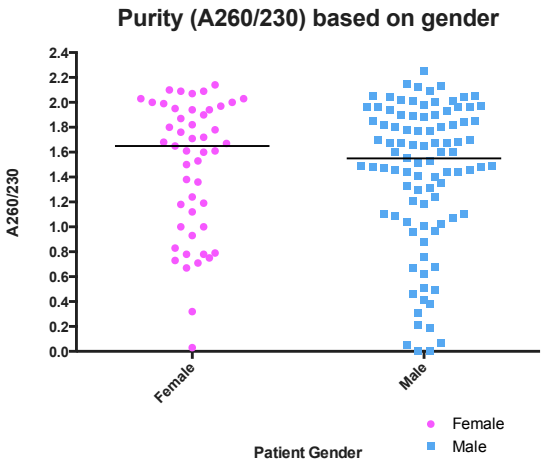
Appendix 3: Yield, purity, correlation and bias in RNA isolation from esophageal tissue

Appendix 3.6 Gender bias

Patient gender does not significantly affect yield or purity. (A) Total RNA yield (µg) separated by gender, Mann-Whitney test, p = 0.5870; (B) Purity (based on absorbance ratio 260nm/230nm), separated by gender, Mann-Whitney test, p = 0.5654.

(A)

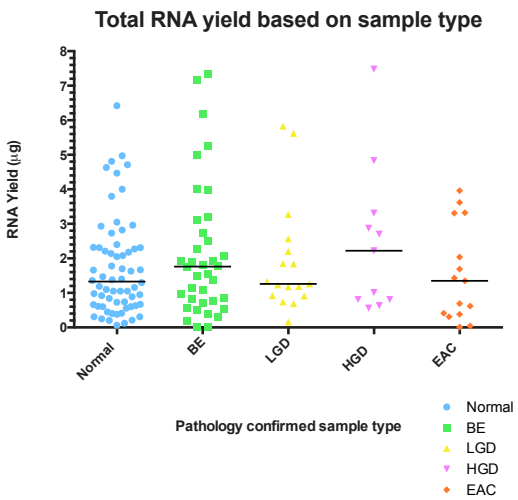
(B)



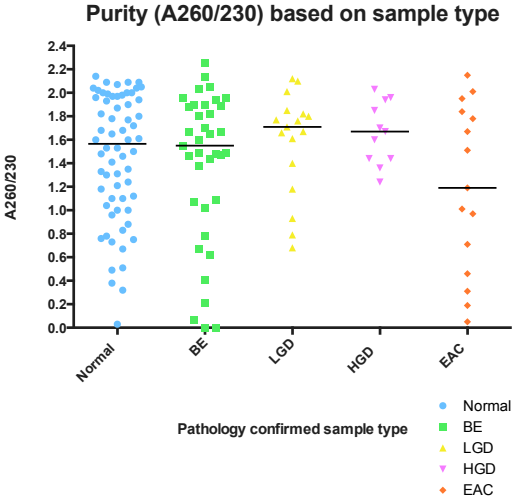
Appendix 3.7 Sample type bias

Sample type does not significantly affect yield or purity. (A) Total RNA yield (µg), separated by pathology confirmed sample type, Kruskal-Wallis test, p = 0.5870; (B) Purity (based on absorbance ratio 260nm/230nm), separated by pathology confirmed sample type, Kruskal-Wallis test, p = 0.5654.

(A)



(B)

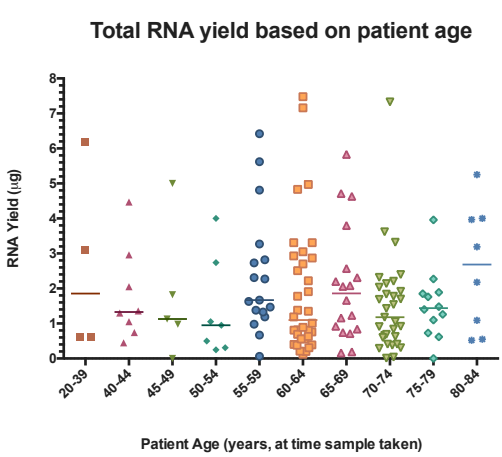


Appendix 3: Yield, purity, correlation and bias in RNA isolation from esophageal tissue

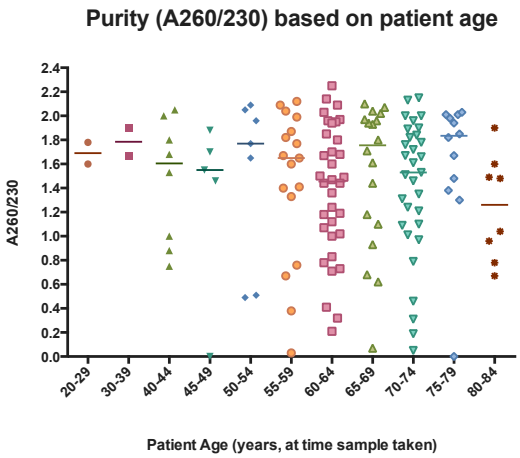
Appendix 3.8 Patient age bias

Patient age (at the time the sample was taken) does not significantly affect yield or purity. (A) Total RNA yield (µg), separated by patient age, Kruskal-Wallis test, p = 0.6571; (B) Purity (based on absorbance ratio 260nm/230nm), separated by patient age, Kruskal-Wallis test, p = 0.5375.

(A)



(B)

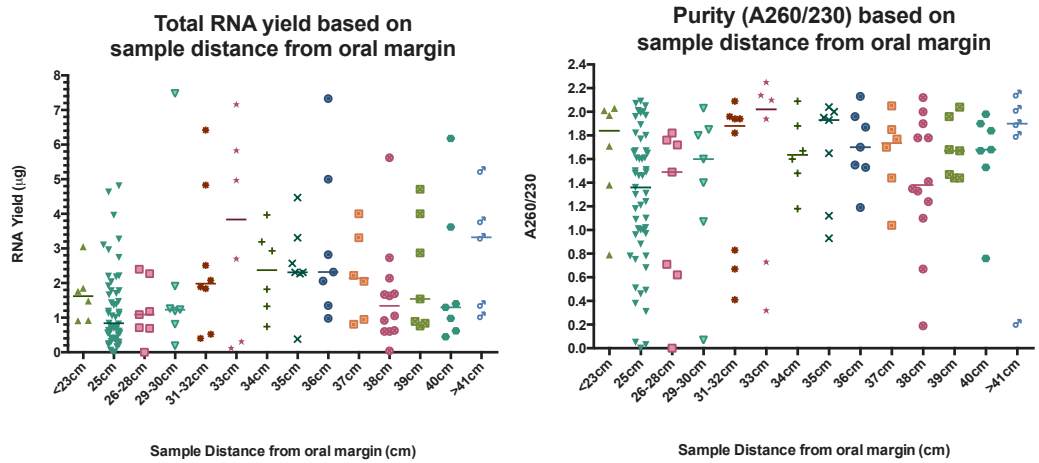


Appendix 3.9 Sample distance from oral margin bias

Sample distance from the oral margin (normal samples typically taken at 25cm; diseased samples distally, dependent on esophageal length) does not significantly affect purity. Sample distance from the oral margin is significantly associated with total RNA yield, however when one-way ANOVA (multiple comparisons) were performed there is no significant difference between the groups when analyzed sequentially. (A) Total RNA yield (µg), separated by sample distance from oral margin, Kruskal-Wallis test, p = 0.0265; (B) Purity (based on absorbance ratio 260nm/230nm), separated by sample distance from oral margin, Kruskal-Wallis test, p = 0.2460.

Appendix 3: Yield, purity, correlation and bias in RNA isolation from esophageal tissue

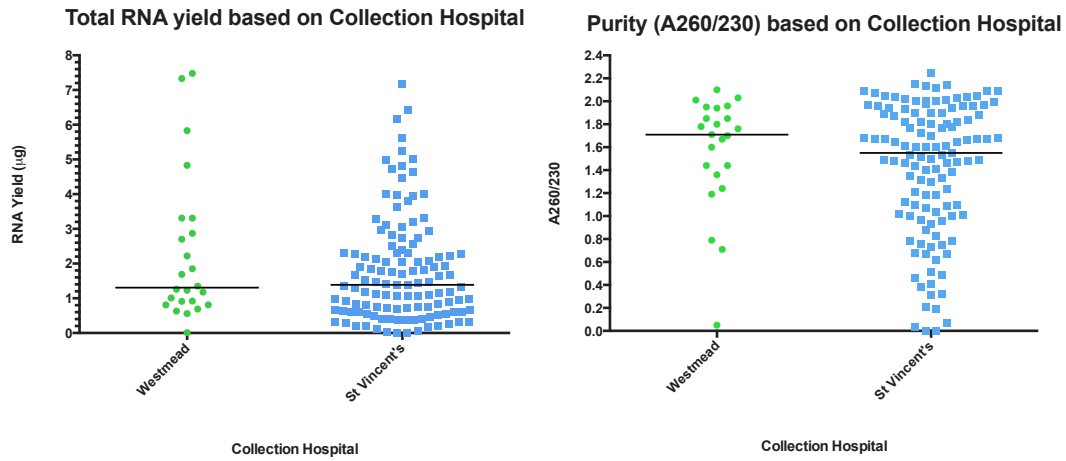
(A) (B)



Appendix 3.10 Collection hospital bias

Collection hospital (standardized protocols) does not significantly affect yield or purity. (A) Total RNA yield (µg) separated by collection hospital, Mann-Whitney test, p = 0.3795; (B) Purity (based on absorbance ratio 260nm/230nm), separated by collection hospital, Mann-Whitney test, p = 0.3265.

(A) (B)



Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

For all results tabulated below, selection criteria pertaining to control tissue (duodenal and proximal stomach) as well as normal peripheral blood were NOT applied. Methylation averages for these specimens are shown for interest only. Additionally, no cut-off for baseline methylation has been applied and as such, may vary. Appendix 4 lists purely the top differentially methylated regions. Chapter 6 (and Appendix 7) pertain to data once biomarker specific selection criteria and cut-offs have been applied.

Table 1: Top 100 hyper- and hypo-methylated probes for differentiating intervention-requiring disease from normal squamous epithelium

55,424 hypermethylated and 20,876 hypomethylated CpG sites were identified as differentially methylated between normal squamous epithelium and intervention requiring disease (N v HGD-EAC, FDR < 0.01 and $\Delta\beta \geq 0.20$). Gencode v19 was used for annotation[204], gene name protein-coding when not specified. logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, β_{base}: average methylation in baseline samples, here, normal squamous epithelium, Δβ: difference in average methylation between comparison groups, here, between HGD-EAC and N, β_{duo}: average methylation in duodenal epithelium, β_{stom}: average methylation in proximal stomach epithelium, β_{blood}: average methylation in peripheral blood from disease-free patients. Top 100 hyper- and hypo-methylated sites based on greatest |Δβ|; listed in order of decreasing |Δβ|.

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
Top 100 hyper-methylated sites from the N v HGD-EAC comparison (differentiating intervention requiring disease from normal squamous epithelium)													
AUTS2	chr7	69064092	69064094	cg17027195	8.5872	20.4221	3.06E-15	1.77E-12	0.0110	0.7995	0.0215	0.0101	0.0110
C8orf34	chr8	69244509	69244511	cg21479226	8.0181	21.2747	1.36E-15	9.92E-13	0.0160	0.7920	0.3088	0.0350	0.0157
N/A	chr1	24612044	24612046	cg10881071	6.1419	19.0719	1.18E-14	4.66E-12	0.1138	0.7869	0.9353	0.9328	0.8912
N/A	chr5	178781562	178781564	cg09068128	6.0657	9.0833	1.10E-08	2.26E-07	0.1200	0.7813	0.9403	0.9486	0.9177
EPHA6	chr3	96532042	96532044	cg11090352	7.3684	14.4170	2.69E-12	2.47E-10	0.0242	0.7795	0.1033	0.0426	0.0111
N/A	chr10	126314679	126314681	cg04426802	6.4090	17.5625	5.98E-14	1.47E-11	0.1755	0.7721	0.9738	0.9621	0.9653
EPB41L3	chr18	5543270	5543272	cg18543270	7.2084	17.9665	3.83E-14	1.07E-11	0.0259	0.7714	0.1119	0.0600	0.0185
N/A	chr6	41554528	41554530	cg15760474	6.1398	15.0013	1.26E-12	1.39E-10	0.1823	0.7579	0.7390	0.5747	0.9176
SCOC	chr4	141295028	141295030	cg21986225	8.2781	15.8192	4.55E-13	6.44E-11	0.0104	0.7550	0.0275	0.0130	0.0100
ADD2	chr2	70995458	70995460	cg15170605	7.1315	20.6494	2.46E-15	1.51E-12	0.0246	0.7550	0.2279	0.0560	0.0191
DOK5	chr20	53091928	53091930	cg19991022	7.7318	15.7062	5.22E-13	7.15E-11	0.0154	0.7536	0.1233	0.0190	0.0100
FOXI2	chr10	129535377	129535379	cg16642284	5.9403	14.8238	1.58E-12	1.65E-10	0.0681	0.7496	0.4257	0.1782	0.1003
DLGAP4	chr20	34894647	34894649	cg03414318	6.8801	19.4492	8.05E-15	3.55E-12	0.0289	0.7492	0.2342	0.0657	0.0448

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
TBCD	chr17	80816850	80816852	cg07769421	6.0985	14.4443	2.59E-12	2.40E-10	0.1992	0.7454	0.9454	0.9180	0.8687
N/A	chr17	74696736	74696738	cg13387113	5.5898	10.0197	2.06E-09	5.40E-08	0.1507	0.7445	0.7985	0.8077	0.9247
MTRR	chr5	7850202	7850204	cg12539796	8.1530	17.7910	4.64E-14	1.22E-11	0.0107	0.7445	0.0427	0.0260	0.0134
N/A	chr17	80794273	80794275	cg18121066	5.7747	15.4522	7.15E-13	9.03E-11	0.0767	0.7431	0.9454	0.9095	0.7214
PIGG	chr4	512993	512995	cg18907098	5.7269	19.9362	4.93E-15	2.50E-12	0.1766	0.7425	0.9431	0.9259	0.9057
ZNF385B	chr2	180726327	180726329	cg06695611	7.7510	16.7092	1.58E-13	2.98E-11	0.0141	0.7412	0.1514	0.0123	0.0115
SLC2A9	chr4	10024411	10024413	cg11124021	5.6511	10.5456	8.42E-10	2.53E-08	0.1755	0.7390	0.9539	0.9484	0.9480
EPB41L3	chr18	5543547	5543549	cg07352438	6.4517	16.3496	2.41E-13	4.03E-11	0.0380	0.7376	0.1180	0.1145	0.0348
BCAT1	chr12	25055966	25055968	cg20399616	6.5540	13.9860	4.78E-12	3.85E-10	0.0348	0.7373	0.5663	0.0695	0.0109
HYDIN	chr16	71264578	71264580	cg09918510	6.4279	10.8229	5.31E-10	1.72E-08	0.0384	0.7364	0.2870	0.0850	0.0128
GSC	chr14	95235124	95235126	cg01163842	7.1154	17.0611	1.05E-13	2.20E-11	0.0222	0.7364	0.1775	0.0669	0.0262
DNAJC6	chr1	65731431	65731433	cg26615127	6.1030	15.3062	8.58E-13	1.03E-10	0.0503	0.7343	0.0546	0.0445	0.0434
ACAN	chr15	89346628	89346630	cg16968596	7.6171	17.5505	6.06E-14	1.48E-11	0.0147	0.7312	0.1983	0.0286	0.0114
FLI1	chr11	128564873	128564875	cg11017065	7.7637	23.7213	1.54E-16	2.13E-13	0.0131	0.7294	0.2579	0.0334	0.0115
N/A	chr8	119626988	119626990	cg05230834	5.4692	15.1480	1.05E-12	1.20E-10	0.1744	0.7290	0.8398	0.8575	0.9124
HMX3	chr10	124895447	124895449	cg18685408	6.9471	12.4102	4.45E-11	2.25E-09	0.0241	0.7290	0.0402	0.0119	0.0108
PRDM2	chr1	14026583	14026585	cg00922376	8.0579	24.6511	7.09E-17	1.28E-13	0.0105	0.7289	0.0157	0.0162	0.0100
N/A	chr4	1215111	1215113	cg12389346	5.6259	18.3971	2.41E-14	7.60E-12	0.1948	0.7279	0.9564	0.9447	0.9617
N/A	chr13	114797348	114797350	cg22906709	5.6538	13.6703	7.34E-12	5.39E-10	0.1978	0.7277	0.8890	0.8004	0.9618
DNM3	chr1	171810321	171810323	cg17154724	8.0061	11.0959	3.41E-10	1.19E-08	0.0107	0.7253	0.3670	0.0246	0.0100
GNAQP1 pseudogene	chr2	132182670	132182672	cg20148575	5.8178	16.3811	2.32E-13	3.92E-11	0.0609	0.7244	0.3326	0.1720	0.0135
ST18	chr8	53322509	53322511	cg27649037	6.0936	14.7908	1.65E-12	1.70E-10	0.2290	0.7240	0.9473	0.9406	0.9548
ZNF790	chr19	37329089	37329091	cg10734240	7.1385	14.4022	2.74E-12	2.50E-10	0.0201	0.7231	0.0260	0.0496	0.0258
N/A	chr19	13112282	13112284	cg22824635	6.6093	18.5472	2.05E-14	6.85E-12	0.2479	0.7220	0.9762	0.9629	0.9821
N/A	chr1	76080726	76080728	cg07923233	7.6957	12.0282	7.90E-11	3.57E-09	0.0132	0.7218	0.0697	0.0283	0.0129
GSG1L	chr16	28074979	28074981	cg00843236	7.5083	23.7918	1.45E-16	2.06E-13	0.0151	0.7215	0.1646	0.0278	0.0131
NTNG1	chr1	107683774	107683776	cg07155336	7.6575	24.6729	6.96E-17	1.26E-13	0.0135	0.7214	0.1629	0.0296	0.0109
BRINP3	chr1	190447289	190447291	cg23010538	5.6633	12.1733	6.34E-11	2.98E-09	0.0692	0.7210	0.2565	0.1019	0.0181
RUNDC3B	chr7	87257537	87257539	cg18542829	7.9486	30.1483	1.20E-18	6.94E-15	0.0109	0.7207	0.0203	0.0148	0.0105
PTGDR	chr14	52734524	52734526	cg05302386	7.4733	12.0664	7.45E-11	3.40E-09	0.0153	0.7191	0.2262	0.0257	0.0111
N/A	chr15	28352097	28352099	cg03061682	7.8673	29.1240	2.43E-18	1.13E-14	0.0114	0.7183	0.2481	0.0151	0.0100
SYT9	chr11	7273147	7273149	cg18560328	7.0229	17.2739	8.26E-14	1.86E-11	0.0214	0.7182	0.0781	0.0428	0.0123
UNC80	chr2	210636349	210636351	cg12903638	6.4373	12.5148	3.81E-11	1.99E-09	0.0337	0.7177	0.1709	0.0465	0.0198

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
TSHZ3	chr19	31841662	31841664	cg11199770	7.2866	17.1392	9.63E-14	2.06E-11	0.0173	0.7158	0.1983	0.0695	0.0134
N/A	chr1	9400774	9400776	cg24597774	5.2908	14.3981	2.76E-12	2.51E-10	0.1803	0.7157	0.8990	0.9162	0.8934
TUB	chr11	8040550	8040552	cg15344220	7.5160	18.8565	1.48E-14	5.38E-12	0.0144	0.7137	0.1100	0.0175	0.0219
SPOCK1	chr5	136834491	136834493	cg14650610	6.4242	20.5092	2.81E-15	1.67E-12	0.0332	0.7135	0.2896	0.0515	0.0173
N/A	chr1	9400738	9400740	cg04778236	6.1980	14.7910	1.65E-12	1.70E-10	0.2469	0.7132	0.9662	0.9599	0.9370
N/A	chr17	66292374	66292376	cg04173852	5.9030	16.2672	2.65E-13	4.32E-11	0.2358	0.7128	0.9297	0.9547	0.9581
FOXB2	chr9	79633736	79633738	cg14487131	6.1901	11.8285	1.07E-10	4.59E-09	0.0399	0.7122	0.1931	0.0912	0.0184
VWC2	chr7	49813087	49813089	cg01893212	6.3335	20.5484	2.71E-15	1.63E-12	0.0354	0.7121	0.4188	0.1283	0.0140
ZIK1	chr19	58095658	58095660	cg26246807	6.1910	11.3235	2.36E-10	8.77E-09	0.0398	0.7120	0.3582	0.0649	0.0129
N/A	chr15	69099677	69099679	cg16068038	5.5744	15.1584	1.03E-12	1.19E-10	0.2186	0.7116	0.9181	0.8627	0.9019
PTPRT	chr20	41818769	41818771	cg17859110	7.3846	22.1261	6.20E-16	5.59E-13	0.0157	0.7116	0.1457	0.0265	0.0130
PRKAR1B	chr7	752291	752293	cg13895235	5.8027	20.2254	3.71E-15	2.02E-12	0.0558	0.7115	0.2135	0.0665	0.0471
TFAP2B	chr6	50791201	50791203	cg27260772	5.4339	11.3418	2.30E-10	8.55E-09	0.0811	0.7112	0.5176	0.1912	0.0563
PTPN5	chr11	18813190	18813192	cg11334818	5.6685	17.1091	9.96E-14	2.11E-11	0.0632	0.7111	0.3172	0.1465	0.0503
N/A	chr10	556547	556549	cg26511507	5.4479	13.9426	5.06E-12	4.03E-10	0.2101	0.7106	0.8678	0.6674	0.8797
NKX2-2	chr20	21492913	21492915	cg22474464	6.4319	13.1029	1.62E-11	1.00E-09	0.0324	0.7105	0.1869	0.0796	0.0211
N/A	chr1	245524537	245524539	cg03130248	5.5861	16.8668	1.32E-13	2.59E-11	0.2213	0.7105	0.9259	0.9109	0.9190
ADHFE1	chr8	67344664	67344666	cg20295442	6.3773	15.0959	1.12E-12	1.27E-10	0.0337	0.7099	0.1321	0.0720	0.0162
ADHFE1	chr8	67344719	67344721	cg20912169	6.9165	14.8556	1.52E-12	1.59E-10	0.0221	0.7095	0.1596	0.0685	0.0140
GDF1 / CERS1	chr19	19007310	19007312	cg03351460	5.9629	16.2050	2.86E-13	4.58E-11	0.0476	0.7094	0.1881	0.1565	0.0601
N/A	chr10	105510712	105510714	cg00730670	5.7834	17.4064	7.12E-14	1.66E-11	0.2350	0.7092	0.9522	0.9635	0.9768
IGDCC3	chr15	65670303	65670305	cg01107006	7.4118	13.6773	7.27E-12	5.34E-10	0.0152	0.7090	0.0340	0.0172	0.0100
N/A	chr1	76081961	76081963	cg27547954	7.8052	17.1992	8.99E-14	1.97E-11	0.0114	0.7086	0.0534	0.0100	0.0111
VGLL3	chr3	87040285	87040287	cg12615137	7.5480	14.4402	2.61E-12	2.41E-10	0.0137	0.7086	0.1509	0.0168	0.0101
LMO1	chr11	8284745	8284747	cg21842523	5.8796	15.3362	8.26E-13	1.01E-10	0.0507	0.7080	0.0649	0.0613	0.0180
CTNND2	chr5	11904113	11904115	cg07195011	6.8203	18.7112	1.73E-14	6.03E-12	0.0235	0.7079	0.2626	0.0710	0.0114
N/A	chr17	78865086	78865088	cg10035831	5.3191	14.8846	1.46E-12	1.55E-10	0.2029	0.7075	0.8951	0.9007	0.8697
VAV3	chr1	108232420	108232422	cg09896211	5.0877	11.4982	1.79E-10	6.98E-09	0.1530	0.7070	0.8691	0.8668	0.8836
C11orf96	chr11	43963906	43963908	cg20062650	7.2485	12.0690	7.42E-11	3.39E-09	0.0169	0.7066	0.0396	0.0433	0.0125
N/A	chr11	20618229	20618231	cg02027945	5.1649	7.1032	5.54E-07	6.62E-06	0.1111	0.7065	0.3202	0.1272	0.0176
N/A	chr13	101194646	101194648	cg20198393	5.6436	17.4991	6.42E-14	1.54E-11	0.2312	0.7065	0.9431	0.9248	0.9199
GDF10	chr10	48438723	48438725	cg04110601	7.4536	18.3007	2.67E-14	8.16E-12	0.0145	0.7061	0.0959	0.0322	0.0164
CALML3	chr10	5567477	5567479	cg20691436	6.4111	13.6095	7.98E-12	5.77E-10	0.2630	0.7051	0.9746	0.9766	0.9472
IRX4	chr5	1883004	1883006	cg07882671	6.5920	13.2667	1.29E-11	8.41E-10	0.0276	0.7050	0.3180	0.1022	0.0205

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
HMX3	chr10	124895473	124895475	cg00560482	6.4198	14.7924	1.65E-12	1.70E-10	0.0316	0.7049	0.0529	0.0374	0.0181
N/A	chr13	101194665	101194667	cg04832557	5.1482	17.5658	5.96E-14	1.46E-11	0.1842	0.7048	0.9134	0.8902	0.8843
RUNDC3B	chr7	87257672	87257674	cg17960051	7.8504	32.0346	3.49E-19	2.87E-15	0.0107	0.7041	0.0110	0.0229	0.0100
N/A	chr7	1850202	1850204	cg26202753	5.1700	15.8041	4.64E-13	6.52E-11	0.1050	0.7036	0.8622	0.6939	0.8688
N/A	chr3	184301729	184301731	cg24000814	5.9076	10.6415	7.17E-10	2.21E-08	0.0477	0.7028	0.5302	0.6924	0.0364
AEBP1	chr7	44143992	44143994	cg14249876	7.3599	22.6307	3.95E-16	4.10E-13	0.0152	0.7025	0.1720	0.0488	0.0106
N/A	chr3	197121409	197121411	cg05485379	6.6472	14.0083	4.63E-12	3.77E-10	0.2723	0.7018	0.9299	0.9498	0.9707
N/A	chr9	135465434	135465436	cg14192291	5.8287	12.7421	2.73E-11	1.52E-09	0.0508	0.7016	0.2437	0.1211	0.0403
RP11-834C11.7 pseudogene	chr12	54473533	54473535	cg26407571	5.3786	12.2952	5.28E-11	2.58E-09	0.2208	0.7010	0.9217	0.3897	0.9488
GPR123	chr10	134901296	134901298	cg15825786	6.4769	17.0681	1.04E-13	2.19E-11	0.0295	0.7010	0.2214	0.0372	0.0217
TMEM178B	chr7	140773904	140773906	cg07028821	7.3338	22.2452	5.57E-16	5.22E-13	0.0154	0.7009	0.0884	0.0460	0.0126
LAMA1	chr18	7116976	7116978	cg22455914	6.5713	15.3173	8.46E-13	1.02E-10	0.0274	0.7007	0.1955	0.0875	0.0249
PLXNA4	chr7	132262352	132262354	cg07258916	7.2773	17.7944	4.63E-14	1.22E-11	0.0160	0.7002	0.0292	0.0272	0.0158
N/A	chr5	510420	510422	cg17778888	5.4888	9.1154	1.04E-08	2.15E-07	0.2306	0.7002	0.9613	0.9541	0.9667
LRP1B	chr2	142887885	142887887	cg20443778	6.1542	10.4797	9.40E-10	2.78E-08	0.0381	0.7001	0.1785	0.0720	0.0457
HS3ST4	chr16	25703527	25703529	cg27014135	7.6340	25.9176	2.58E-17	6.02E-14	0.0123	0.7000	0.0603	0.0100	0.0100
CTD-2245F17.3 lincRNA	chr19	53700526	53700528	cg23021477	7.2922	19.1884	1.05E-14	4.30E-12	0.0158	0.6997	0.0142	0.0530	0.0149
PURG	chr8	30890619	30890621	cg18324126	7.5802	16.6990	1.60E-13	3.00E-11	0.0127	0.6990	0.0399	0.0179	0.0100
DNM3	chr1	171810299	171810301	cg27429080	7.3266	12.1430	6.64E-11	3.10E-09	0.0153	0.6989	0.4187	0.0832	0.0129
N/A	chr7	20817529	20817531	cg08650910	5.6553	16.0507	3.44E-13	5.23E-11	0.0582	0.6986	0.2162	0.0689	0.0400
Top 100 hypo-methylated sites from the N v HGD-EAC comparison (differentiating intervention requiring disease from normal squamous epithelium)													
SSBP3	chr1	54822030	54822032	cg00172603	-5.6739	-14.3153	3.07E-12	2.73E-10	0.8606	-0.7527	0.2791	0.2337	0.0972
N/A	chr19	2278617	2278619	cg27341866	-5.7136	-18.3377	2.57E-14	7.91E-12	0.9259	-0.7336	0.2126	0.8822	0.8772
N/A	chr21	36577538	36577540	cg09200260	-5.0717	-14.2823	3.21E-12	2.84E-10	0.8186	-0.7003	0.5718	0.1119	0.0756
N/A	chr10	70821370	70821372	cg07133930	-5.0344	-11.2392	2.70E-10	9.83E-09	0.8136	-0.6961	0.3472	0.6775	0.3372
N/A	chr14	107253272	107253274	cg19901523	-4.8798	-10.6762	6.77E-10	2.11E-08	0.8391	-0.6886	0.6182	0.7181	0.0336
RNF126P1 pseudogene	chr17	55122827	55122829	cg02486253	-4.8344	-5.3735	2.58E-05	0.000179947	0.8448	-0.6846	0.2897	0.1549	0.0830
N/A	chr4	186434992	186434994	cg25581330	-5.6293	-15.4107	7.53E-13	9.38E-11	0.7379	-0.6841	0.1176	0.1204	0.0613
LAIR1	chr19	54876727	54876729	cg21878746	-4.8264	-6.7891	1.08E-06	1.18E-05	0.8396	-0.6838	0.4442	0.7485	0.0554
N/A	chr13	114880888	114880890	cg15972148	-4.8456	-15.3501	8.12E-13	9.92E-11	0.8762	-0.6787	0.4216	0.6814	0.5137
N/A	chr3	169384409	169384411	cg04120686	-4.8181	-18.1167	3.26E-14	9.46E-12	0.8055	-0.6775	0.4131	0.1603	0.0521

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
LAIR1	chr19	54876748	54876750	cg01515802	-4.8996	-6.5208	1.94E-06	1.95E-05	0.8897	-0.6770	0.5165	0.7837	0.0349
N/A	chr11	69264656	69264658	cg26540315	-4.9251	-11.8641	1.01E-10	4.39E-09	0.7691	-0.6703	0.3372	0.1772	0.1022
CAPN2	chr1	223936798	223936800	cg06756211	-5.0636	-5.6882	1.25E-05	9.67E-05	0.7513	-0.6685	0.0895	0.1809	0.0569
ETS1	chr11	128391428	128391430	cg26503877	-4.9611	-7.0405	6.33E-07	7.43E-06	0.7486	-0.6613	0.5437	0.4361	0.1129
N/A	chr7	151442350	151442352	cg06436185	-4.5894	-6.5289	1.90E-06	1.92E-05	0.8375	-0.6612	0.4179	0.3406	0.4143
N/A	chr1	225942841	225942843	cg14520947	-4.6784	-20.4388	3.01E-15	1.76E-12	0.8759	-0.6598	0.3817	0.3076	0.4668
N/A	chr5	76034886	76034888	cg24722577	-4.6761	-14.3522	2.93E-12	2.63E-10	0.7668	-0.6528	0.0993	0.3956	0.0785
MSMB	chr10	51549504	51549506	cg10326726	-4.6774	-13.1745	1.47E-11	9.31E-10	0.8908	-0.6491	0.2547	0.2317	0.8917
N/A	chr21	36577637	36577639	cg21172011	-4.9014	-14.4596	2.54E-12	2.36E-10	0.7325	-0.6486	0.2853	0.0789	0.1063
SERPINA11	chr14	94917778	94917780	cg04368971	-4.5131	-7.2304	4.24E-07	5.25E-06	0.7775	-0.6448	0.2661	0.3842	0.3291
LPIN1	chr2	11917786	11917788	cg24011073	-4.8677	-7.4232	2.84E-07	3.71E-06	0.9174	-0.6419	0.6594	0.8887	0.2334
N/A	chr1	8143890	8143892	cg19274341	-4.4702	-11.7495	1.21E-10	5.07E-09	0.7792	-0.6418	0.2164	0.2048	0.3203
N/A	chr20	48998700	48998702	cg20430841	-4.3960	-18.9305	1.37E-14	5.13E-12	0.8106	-0.6417	0.2717	0.4300	0.1799
N/A	chr2	127955155	127955157	cg24923516	-4.4270	-14.5283	2.32E-12	2.21E-10	0.8578	-0.6388	0.4572	0.8108	0.3561
N/A	chr3	51104050	51104052	cg13302785	-4.3935	-11.7941	1.13E-10	4.79E-09	0.7812	-0.6360	0.4936	0.6309	0.2261
GDF15	chr19	18497142	18497144	cg12008047	-5.4483	-13.1243	1.58E-11	9.83E-10	0.9535	-0.6338	0.3536	0.3217	0.9402
N/A	chr8	108208343	108208345	cg21186098	-4.4615	-10.9903	4.04E-10	1.37E-08	0.7510	-0.6306	0.1996	0.4329	0.2910
N/A	chr5	123985367	123985369	cg21213593	-4.4288	-16.3170	2.50E-13	4.14E-11	0.7521	-0.6286	0.2156	0.2481	0.0580
RP11-443B7.2 lincRNA	chr1	235091186	235091188	cg05926640	-4.4112	-17.6811	5.24E-14	1.34E-11	0.7500	-0.6264	0.1799	0.3767	0.2003
RP11-443B7.2 lincRNA	chr1	235091105	235091107	cg20155035	-4.2403	-18.6750	1.79E-14	6.21E-12	0.7972	-0.6250	0.1765	0.3989	0.2108
IL18	chr11	112035944	112035946	cg05687149	-4.2094	-15.8409	4.43E-13	6.31E-11	0.8364	-0.6199	0.4286	0.2885	0.2485
KRT7	chr12	52627437	52627439	cg22958090	-4.4639	-6.0869	5.06E-06	4.46E-05	0.8921	-0.6196	0.8111	0.8005	0.2288
N/A	chr1	223936811	223936813	cg19598416	-4.3328	-5.0767	5.14E-05	0.000324477	0.7408	-0.6166	0.1504	0.2352	0.0986
SERPINA11	chr14	94917626	94917628	cg21007342	-4.2892	-16.8636	1.32E-13	2.60E-11	0.7461	-0.6154	0.1383	0.1462	0.2459
N/A	chr3	169384518	169384520	cg14580747	-4.3049	-14.5044	2.40E-12	2.26E-10	0.7420	-0.6150	0.3788	0.1520	0.0663
N/A	chr16	85981719	85981721	cg01574513	-4.1659	-7.0097	6.76E-07	7.86E-06	0.7753	-0.6141	0.4753	0.3828	0.0759
N/A	chr7	98311381	98311383	cg02928476	-4.4793	-10.0191	2.07E-09	5.41E-08	0.7149	-0.6139	0.5553	0.6538	0.0290
RP11-432J24.2 lincRNA	chr10	134231548	134231550	cg23249922	-4.2695	-12.7922	2.54E-11	1.44E-09	0.8710	-0.6117	0.5497	0.3475	0.4207
N/A	chr2	208631683	208631685	cg15714846	-4.0914	-13.3042	1.22E-11	8.07E-10	0.8193	-0.6092	0.2204	0.2493	0.1491
N/A	chr20	22562477	22562479	cg20504791	-4.1481	-8.1457	6.61E-08	1.05E-06	0.7611	-0.6088	0.5212	0.0951	0.1992
N/A	chr7	116232818	116232820	cg10087556	-4.2184	-16.4325	2.18E-13	3.76E-11	0.7388	-0.6069	0.1817	0.2355	0.0898

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
N/A	chr8	41176849	41176851	cg19656415	-4.2337	-15.3673	7.95E-13	9.79E-11	0.7348	-0.6064	0.1796	0.2020	0.1455
N/A	chr8	25059251	25059253	cg00157012	-4.0847	-16.2643	2.66E-13	4.34E-11	0.8336	-0.6057	0.3308	0.3344	0.2122
N/A	chr16	83869926	83869928	cg09683440	-4.1157	-9.4045	6.14E-09	1.36E-07	0.7564	-0.6045	0.3438	0.1502	0.0694
MX2	chr21	42734265	42734267	cg21130374	-4.5215	-13.7511	6.57E-12	4.93E-10	0.6897	-0.6014	0.1585	0.3620	0.0460
N/A	chr4	4397941	4397943	cg09852389	-4.2394	-13.0724	1.70E-11	1.04E-09	0.7226	-0.6014	0.3718	0.3827	0.1824
N/A	chr2	239403002	239403004	cg03217587	-4.0096	-12.7583	2.67E-11	1.49E-09	0.8216	-0.5992	0.2144	0.6836	0.0663
N/A	chr3	72870837	72870839	cg18720486	-3.9874	-17.0960	1.01E-13	2.13E-11	0.8057	-0.5984	0.3567	0.3511	0.2215
N/A	chr11	268949	268951	cg13185005	-4.1682	-15.3547	8.07E-13	9.89E-11	0.7279	-0.5984	0.2285	0.4909	0.0524
N/A	chr4	184949986	184949988	cg00331101	-3.9687	-10.1291	1.71E-09	4.60E-08	0.8020	-0.5965	0.5905	0.6668	0.0826
N/A	chr22	47082259	47082261	cg06766034	-3.9807	-12.5864	3.43E-11	1.82E-09	0.8215	-0.5958	0.4129	0.2894	0.3885
N/A	chr19	10715611	10715613	cg06943912	-3.9608	-10.4937	9.18E-10	2.72E-08	0.7926	-0.5955	0.5468	0.6639	0.1094
SEMA3C	chr7	80549170	80549172	cg09225287	-3.9520	-18.1146	3.26E-14	9.47E-12	0.7952	-0.5946	0.3246	0.6463	0.3443
N/A	chr1	15062528	15062530	cg19626253	-3.9560	-12.9783	1.94E-11	1.16E-09	0.8135	-0.5941	0.2851	0.4036	0.1144
N/A	chr13	114770147	114770149	cg21422164	-5.0308	-10.1190	1.74E-09	4.67E-08	0.9469	-0.5940	0.9426	0.9410	0.8676
N/A	chr16	81328088	81328090	cg00602295	-3.9673	-10.1694	1.59E-09	4.33E-08	0.7645	-0.5926	0.1966	0.1858	0.3247
N/A	chr4	184642693	184642695	cg22078781	-3.9587	-8.5828	2.83E-08	5.05E-07	0.7664	-0.5922	0.2077	0.1905	0.1969
N/A	chr6	1601508	1601510	cg16987638	-3.9510	-10.0682	1.90E-09	5.03E-08	0.7639	-0.5909	0.1608	0.1682	0.2348
N/A	chr3	124711100	124711102	cg19533443	-4.5911	-17.3034	7.99E-14	1.81E-11	0.9237	-0.5893	0.3652	0.5152	0.8713
N/A	chr1	111189870	111189872	cg03257417	-4.1879	-7.6364	1.83E-07	2.54E-06	0.7055	-0.5893	0.3511	0.1311	0.3610
N/A	chr10	115141826	115141828	cg01246622	-4.3119	-11.1481	3.13E-10	1.11E-08	0.6879	-0.5880	0.4296	0.4565	0.0668
N/A	chr11	655578	655580	cg17005319	-5.2175	-17.0047	1.12E-13	2.30E-11	0.9559	-0.5878	0.4669	0.3811	0.9707
SSBP3	chr1	54822502	54822504	cg08641990	-3.9414	-9.7606	3.25E-09	7.94E-08	0.8351	-0.5872	0.4254	0.3655	0.2639
CCDC141	chr2	179914863	179914865	cg24335070	-4.0935	-16.1103	3.20E-13	4.97E-11	0.7153	-0.5870	0.1045	0.0773	0.0773
NHLH2	chr1	116383237	116383239	cg23202177	-3.8841	-15.6275	5.75E-13	7.67E-11	0.7912	-0.5870	0.2221	0.1998	0.2623
N/A	chr4	124521390	124521392	cg03862787	-4.3771	-18.5209	2.11E-14	6.97E-12	0.6778	-0.5859	0.1423	0.1150	0.0670
N/A	chr15	70740428	70740430	cg19580937	-3.8822	-7.5942	2.00E-07	2.74E-06	0.7702	-0.5850	0.2969	0.4431	0.1385
N/A	chr3	193973395	193973397	cg21130113	-3.9426	-11.8567	1.03E-10	4.43E-09	0.7411	-0.5841	0.2023	0.2832	0.0435
EGFL7	chr9	139553209	139553211	cg14143326	-3.8708	-14.7132	1.83E-12	1.83E-10	0.8117	-0.5841	0.3177	0.5117	0.1435
N/A	chr17	63225019	63225021	cg04770088	-3.9073	-12.1449	6.62E-11	3.09E-09	0.7470	-0.5826	0.2023	0.2554	0.0615
N/A	chr11	269374	269376	cg11546385	-3.9764	-10.9091	4.61E-10	1.53E-08	0.7277	-0.5826	0.4778	0.6600	0.0421
N/A	chr1	204256845	204256847	cg25407979	-4.1145	-11.0370	3.75E-10	1.29E-08	0.8802	-0.5823	0.5254	0.2149	0.1641
PHACTR1	chr6	13274150	13274152	cg06879394	-4.0535	-6.2205	3.76E-06	3.45E-05	0.7103	-0.5816	0.6272	0.4093	0.0188
N/A	chr6	170532632	170532634	cg24454932	-4.0455	-6.6485	1.47E-06	1.53E-05	0.7110	-0.5813	0.1161	0.4979	0.1287
PRR26	chr10	696331	696333	cg14121234	-6.2385	-16.9982	1.13E-13	2.32E-11	0.9806	-0.5791	0.4200	0.3788	0.9881

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
N/A	chr9	27379517	27379519	cg13648550	-4.0224	-17.9131	4.06E-14	1.11E-11	0.7094	-0.5788	0.1821	0.2131	0.0903
N/A	chr3	193587779	193587781	cg13323489	-4.3353	-6.3166	3.04E-06	2.87E-05	0.6700	-0.5786	0.4743	0.5284	0.0323
N/A	chr6	47468192	47468194	cg19989043	-3.8567	-15.1812	1.00E-12	1.17E-10	0.8295	-0.5781	0.3706	0.3302	0.3069
N/A	chr5	10485606	10485608	cg18430895	-4.4727	-13.1677	1.48E-11	9.37E-10	0.6571	-0.5776	0.2582	0.1692	0.2016
N/A	chr13	113584106	113584108	cg24028634	-3.8155	-4.6509	0.000140331	0.000764349	0.8132	-0.5770	0.2847	0.6844	0.0678
N/A	chr17	16200580	16200582	cg04592811	-3.8622	-19.2255	1.01E-14	4.18E-12	0.8374	-0.5759	0.3949	0.3322	0.4376
N/A	chr5	43039591	43039593	cg04122815	-4.5932	-11.6670	1.38E-10	5.63E-09	0.6458	-0.5755	0.1972	0.3784	0.0495
TBXAS1	chr7	139528854	139528856	cg05711445	-4.0919	-17.2591	8.40E-14	1.88E-11	0.6911	-0.5751	0.2658	0.2005	0.1160
N/A	chr3	112903053	112903055	cg00584840	-3.8988	-9.7576	3.26E-09	7.97E-08	0.8495	-0.5750	0.2711	0.1339	0.9054
N/A	chr10	30316431	30316433	cg12786570	-3.9153	-8.6504	2.49E-08	4.53E-07	0.7191	-0.5740	0.5034	0.5673	0.0321
N/A	chr14	78107676	78107678	cg18674234	-3.8233	-11.3111	2.41E-10	8.92E-09	0.8299	-0.5736	0.5170	0.7104	0.4324
DAPK1	chr9	90114754	90114756	cg13752933	-4.0413	-16.7465	1.51E-13	2.88E-11	0.6945	-0.5732	0.1363	0.1353	0.1725
N/A	chr1	3139866	3139868	cg15727188	-4.0622	-9.2256	8.50E-09	1.80E-07	0.8821	-0.5728	0.2951	0.2725	0.8599
GPSM3	chr6	32164800	32164802	cg06023661	-3.9873	-8.0012	8.80E-08	1.35E-06	0.7021	-0.5727	0.1612	0.1199	0.3821
N/A	chr1	203293432	203293434	cg06637812	-3.8036	-10.3499	1.17E-09	3.35E-08	0.8249	-0.5726	0.1893	0.3542	0.0795
N/A	chr3	169377817	169377819	cg00024967	-3.7569	-8.0374	8.19E-08	1.26E-06	0.7726	-0.5718	0.3621	0.2446	0.0996
N/A	chr15	78726575	78726577	cg20117256	-3.7682	-11.6184	1.48E-10	5.99E-09	0.7573	-0.5710	0.2829	0.4099	0.2888
SDE2	chr1	226187875	226187877	cg13556548	-3.7316	-10.1116	1.76E-09	4.72E-08	0.7984	-0.5687	0.5740	0.2162	0.0802
RP11-627G18.2 antisense	chr18	19757467	19757469	cg15424989	-3.7655	-10.4575	9.76E-10	2.87E-08	0.8219	-0.5685	0.3539	0.2847	0.8355
RIN2	chr20	19870216	19870218	cg06856720	-3.8881	-6.8131	1.03E-06	1.13E-05	0.7097	-0.5680	0.1839	0.5960	0.2525
N/A	chr10	3466851	3466853	cg18692070	-3.7497	-11.0736	3.53E-10	1.23E-08	0.7470	-0.5670	0.5438	0.6523	0.2476
CCRL2	chr3	46448578	46448580	cg19850333	-4.5527	-18.0406	3.54E-14	1.01E-11	0.6364	-0.5670	0.1085	0.1340	0.0762
N/A	chr7	79964737	79964739	cg22110839	-3.7217	-13.6065	8.02E-12	5.78E-10	0.7619	-0.5667	0.1472	0.1527	0.5017
N/A	chr2	8849961	8849963	cg10900049	-5.9226	-20.6386	2.48E-15	1.53E-12	0.9769	-0.5665	0.3620	0.9587	0.9695
CACNA2D3	chr3	55040616	55040618	cg10791541	-3.7253	-13.3777	1.10E-11	7.44E-10	0.8117	-0.5659	0.3341	0.2476	0.2095

Table 2: Top 100 hyper- and hypo-methylated probes for differentiating Barrett's esophagus from normal squamous epithelium
44,201 hypermethylated and 26,985 hypomethylated CpG sites were identified as differentially methylated between normal squamous epithelium and Barrett's esophagus (N v BE, FDR < 0.01 and Δβ ≥ 0.20). Gencode v19 was used for annotation[204], gene name protein-coding when not specified. logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, β_{base}: average methylation in baseline samples, here, normal squamous epithelium, Δβ: difference in average methylation between comparison groups, here, between BE and N, β_{duo}:

average methylation in duodenal epithelium, β_{stom}: average methylation in proximal stomach epithelium, β_{blood}: average methylation in peripheral blood from disease-free patients. Top 100 hyper- and hypo-methylated sites based on greatest |Δβ|; listed in order of decreasing |Δβ|.

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
Top 100 hyper-methylated sites from the N v BE comparison (differentiating Barrett's esophagus from normal squamous epithelium)													
N/A	chr17	80794273	80794275	cg18121066	7.1119	19.0306	1.24E-14	4.54E-12	0.0767	0.8432	0.9454	0.9095	0.7214
N/A	chr5	178781562	178781564	cg09068128	6.9073	10.3436	1.18E-09	3.37E-08	0.1200	0.8224	0.9403	0.9486	0.9177
C8orf34	chr8	69244509	69244511	cg21479226	8.3115	22.0532	6.62E-16	5.81E-13	0.0160	0.8216	0.3088	0.0350	0.0157
N/A	chr1	24612044	24612046	cg10881071	6.6713	20.7157	2.30E-15	1.44E-12	0.1138	0.8152	0.9353	0.9328	0.8912
N/A	chr17	40823929	40823931	cg19987356	6.3464	9.9307	2.41E-09	6.15E-08	0.1166	0.7982	0.8963	0.8786	0.7327
N/A	chr10	105517763	105517765	cg06888746	6.6578	12.2169	5.94E-11	2.83E-09	0.0495	0.7907	0.8187	0.8401	0.0378
SALL1	chr16	51185059	51185061	cg08439930	6.4480	15.4485	7.18E-13	8.83E-11	0.0576	0.7846	0.2432	0.0922	0.0335
TSHZ3	chr19	31841662	31841664	cg11199770	7.8261	18.4082	2.38E-14	7.21E-12	0.0173	0.7824	0.1983	0.0695	0.0134
WDR64	chr1	241912763	241912765	cg05940691	6.3484	10.5502	8.35E-10	2.52E-08	0.1566	0.7814	0.8514	0.9470	0.9402
SLC2A9	chr4	10024411	10024413	cg11124021	6.6280	12.3686	4.73E-11	2.35E-09	0.1755	0.7791	0.9539	0.9484	0.9480
EPB41L3	chr18	5543270	5543272	cg18543270	7.2356	18.0344	3.56E-14	9.72E-12	0.0259	0.7744	0.1119	0.0600	0.0185
N/A	chr14	91862863	91862865	cg10928544	5.9615	15.0008	1.26E-12	1.36E-10	0.1301	0.7730	0.8658	0.9018	0.5149
N/A	chr10	126314679	126314681	cg04426802	6.3651	17.4420	6.84E-14	1.55E-11	0.1755	0.7706	0.9738	0.9621	0.9653
TDRP	chr8	495816	495818	cg19019537	6.7148	15.5632	6.23E-13	7.95E-11	0.0359	0.7605	0.2930	0.0727	0.0150
FOXI2	chr10	129535377	129535379	cg16642284	6.0130	15.0051	1.25E-12	1.35E-10	0.0681	0.7570	0.4257	0.1782	0.1003
HOXD8	chr2	176994763	176994765	cg24416513	6.0974	15.1678	1.02E-12	1.16E-10	0.0609	0.7553	0.3199	0.2774	0.0307
FOXB2	chr9	79633736	79633738	cg14487131	6.5333	12.4844	3.98E-11	2.05E-09	0.0399	0.7538	0.1931	0.0912	0.0184
PIGG	chr4	512993	512995	cg18907098	5.9529	20.7228	2.29E-15	1.43E-12	0.1766	0.7534	0.9431	0.9259	0.9057
GNAQP1 pseudogene	chr2	132182670	132182672	cg20148575	6.0756	17.1069	9.99E-14	2.05E-11	0.0609	0.7530	0.3326	0.1720	0.0135
N/A	chr7	653308	653310	cg18367631	5.9670	14.7145	1.82E-12	1.80E-10	0.1841	0.7497	0.8998	0.9221	0.5078
TBCD	chr17	80816850	80816852	cg07769421	6.0932	14.4317	2.64E-12	2.38E-10	0.1992	0.7452	0.9454	0.9180	0.8687
N/A	chr6	41554528	41554530	cg15760474	5.7845	14.1333	3.92E-12	3.23E-10	0.1823	0.7424	0.7390	0.5747	0.9176
FLI1	chr11	128564873	128564875	cg11017065	7.8597	24.0145	1.20E-16	1.78E-13	0.0131	0.7419	0.2579	0.0334	0.0115
SLITRK5	chr13	88324569	88324571	cg15778745	6.4861	15.7254	5.10E-13	6.87E-11	0.0379	0.7416	0.2914	0.0811	0.0176
PIEZO2	chr18	11149434	11149436	cg03117976	6.6372	12.4613	4.12E-11	2.11E-09	0.0334	0.7413	0.3945	0.0994	0.0160
LAMA1	chr18	7116976	7116978	cg22455914	6.8693	16.0119	3.60E-13	5.32E-11	0.0274	0.7396	0.1955	0.0875	0.0249
N/A	chr4	1215111	1215113	cg12389346	5.8660	19.1821	1.06E-14	4.11E-12	0.1948	0.7390	0.9564	0.9447	0.9617
N/A	chr4	779479	779481	cg17128947	5.7446	15.5642	6.22E-13	7.94E-11	0.0758	0.7390	0.3808	0.6702	0.0610
VAV3	chr1	108232420	108232422	cg09896211	5.5049	12.4412	4.25E-11	2.16E-09	0.1530	0.7383	0.8691	0.8668	0.8836
TFAP2B	chr6	50791201	50791203	cg27260772	5.6826	11.8610	1.02E-10	4.39E-09	0.0811	0.7382	0.5176	0.1912	0.0563

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
ZNF667	chr19	56989542	56989544	cg08063125	7.2364	17.9820	3.77E-14	1.01E-11	0.0204	0.7380	0.3708	0.0690	0.0113
EPB41L3	chr18	5543547	5543549	cg07352438	6.4525	16.3518	2.40E-13	3.93E-11	0.0380	0.7377	0.1180	0.1145	0.0348
N/A	chr1	2084518	2084520	cg04347414	5.4229	14.5112	2.37E-12	2.20E-10	0.1335	0.7351	0.7488	0.8300	0.5873
N/A	chr8	119626988	119626990	cg05230834	5.5565	15.3898	7.73E-13	9.36E-11	0.1744	0.7342	0.8398	0.8575	0.9124
VWC2	chr7	49813087	49813089	cg01893212	6.4982	21.0824	1.63E-15	1.13E-12	0.0354	0.7330	0.4188	0.1283	0.0140
PIEZO2	chr18	11149469	11149471	cg03686593	5.6330	16.0468	3.45E-13	5.17E-11	0.0804	0.7323	0.4569	0.1484	0.0386
N/A	chr1	245524537	245524539	cg03130248	6.1507	18.5717	2.00E-14	6.43E-12	0.2213	0.7315	0.9259	0.9109	0.9190
EPHA6	chr3	96533510	96533512	cg11410023	5.7956	17.5423	6.12E-14	1.43E-11	0.0655	0.7302	0.3008	0.1791	0.0692
RP11-388M20.9 lincRNA	chr16	31238216	31238218	cg03016934	6.1036	14.3832	2.81E-12	2.51E-10	0.2216	0.7298	0.9605	0.9583	0.8788
CTNND2	chr5	11904113	11904115	cg07195011	6.9846	19.1621	1.08E-14	4.14E-12	0.0235	0.7297	0.2626	0.0710	0.0114
DLGAP4	chr20	34894647	34894649	cg03414318	6.7104	18.9695	1.32E-14	4.73E-12	0.0289	0.7282	0.2342	0.0657	0.0448
BRINP3	chr1	190447289	190447291	cg23010538	5.7239	12.3037	5.21E-11	2.55E-09	0.0692	0.7279	0.2565	0.1019	0.0181
SHISA9	chr16	12995295	12995297	cg04342955	6.4946	19.8316	5.48E-15	2.61E-12	0.0343	0.7279	0.2794	0.1079	0.0675
TDRP	chr8	494720	494722	cg03294066	6.8126	16.5778	1.84E-13	3.23E-11	0.0265	0.7271	0.1823	0.0509	0.0180
GNAO1	chr16	56228466	56228468	cg07700514	7.8979	32.0826	3.38E-19	2.64E-15	0.0117	0.7264	0.1532	0.0475	0.0101
LHX1	chr17	35294480	35294482	cg10043865	7.1616	20.7725	2.18E-15	1.38E-12	0.0201	0.7257	0.1969	0.0583	0.0109
N/A	chr19	13112282	13112284	cg22824635	6.7797	19.0252	1.24E-14	4.56E-12	0.2479	0.7253	0.9762	0.9629	0.9821
HS3ST4	chr16	25703527	25703529	cg27014135	7.8061	26.5019	1.64E-17	4.35E-14	0.0123	0.7238	0.0603	0.0100	0.0100
RIMS2	chr8	104512857	104512859	cg01566592	6.7704	17.9758	3.79E-14	1.01E-11	0.0266	0.7224	0.1357	0.0698	0.0137
LINC00404 lincRNA	chr13	112759892	112759894	cg25570913	5.9567	11.7827	1.15E-10	4.86E-09	0.0523	0.7218	0.3602	0.0814	0.0752
N/A	chr17	66292374	66292376	cg04173852	6.1927	17.0656	1.05E-13	2.13E-11	0.2358	0.7218	0.9297	0.9547	0.9581
N/A	chr13	101194665	101194667	cg04832557	5.4139	18.4726	2.22E-14	6.90E-12	0.1842	0.7217	0.9134	0.8902	0.8843
VWC2	chr7	49813032	49813034	cg04904331	6.6846	19.1678	1.07E-14	4.14E-12	0.0283	0.7214	0.3670	0.0832	0.0139
N/A	chr10	105510712	105510714	cg00730670	6.1544	18.5232	2.11E-14	6.69E-12	0.2350	0.7213	0.9522	0.9635	0.9768
N/A	chr1	9400774	9400776	cg24597774	5.3676	14.6072	2.09E-12	2.00E-10	0.1803	0.7205	0.8990	0.9162	0.8934
ST18	chr8	53322509	53322511	cg27649037	5.9832	14.5230	2.34E-12	2.17E-10	0.2290	0.7205	0.9473	0.9406	0.9548
ZNF471	chr19	57019068	57019070	cg00674365	5.6520	12.8609	2.30E-11	1.32E-09	0.0696	0.7203	0.3719	0.1561	0.0312
N/A	chr15	69099677	69099679	cg16068038	5.7739	15.7007	5.26E-13	7.03E-11	0.2186	0.7201	0.9181	0.8627	0.9019
N/A	chr1	15789861	15789863	cg04503318	5.3179	14.4895	2.44E-12	2.25E-10	0.1754	0.7192	0.8649	0.8397	0.9246
PTPRT	chr20	41818769	41818771	cg17859110	7.4392	22.2896	5.35E-16	4.93E-13	0.0157	0.7190	0.1457	0.0265	0.0130
N/A	chr17	80744832	80744834	cg11524039	5.3585	16.1036	3.23E-13	4.91E-11	0.1837	0.7186	0.8990	0.9007	0.9108
N/A	chr11	20618229	20618231	cg02027945	5.2832	7.2659	3.94E-07	4.94E-06	0.1111	0.7184	0.3202	0.1272	0.0176
C8orf34	chr8	69243485	69243487	cg22001496	6.6837	15.0652	1.16E-12	1.28E-10	0.0278	0.7183	0.3725	0.1192	0.0202

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
POU3F3	chr2	105470560	105470562	cg19513834	5.9548	10.9242	4.50E-10	1.50E-08	0.0509	0.7178	0.2020	0.1018	0.0115
ZFP28	chr19	57050358	57050360	cg24152605	6.8110	23.1491	2.51E-16	2.91E-13	0.0250	0.7174	0.2269	0.0704	0.0238
ADCY4	chr14	24803872	24803874	cg23179456	6.7146	26.0050	2.41E-17	5.72E-14	0.0269	0.7168	0.1412	0.2256	0.0147
TBCD	chr17	80816701	80816703	cg23892310	5.5093	12.0764	7.34E-11	3.35E-09	0.2051	0.7165	0.9434	0.9207	0.8632
IRX4	chr5	1883004	1883006	cg07882671	6.6770	13.4378	1.01E-11	6.85E-10	0.0276	0.7164	0.3180	0.1022	0.0205
PAX9	chr14	37128585	37128587	cg04415798	5.9812	12.3324	4.99E-11	2.46E-09	0.0489	0.7157	0.3165	0.1758	0.0629
MTMR7	chr8	17271066	17271068	cg12296772	5.4776	11.2678	2.58E-10	9.51E-09	0.0797	0.7145	0.3160	0.2074	0.0862
SLC6A15	chr12	85306915	85306917	cg03064067	5.5052	15.3939	7.69E-13	9.31E-11	0.0770	0.7142	0.3987	0.2046	0.2571
BCAT1	chr12	25055966	25055968	cg20399616	6.3647	13.5819	8.29E-12	5.85E-10	0.0348	0.7134	0.5663	0.0695	0.0109
TRIM71	chr3	32860466	32860468	cg18249634	5.8406	9.9269	2.43E-09	6.19E-08	0.0545	0.7132	0.2975	0.0849	0.1064
N/A	chr2	468178	468180	cg13699355	6.4120	23.6609	1.62E-16	2.17E-13	0.0333	0.7125	0.3016	0.1055	0.0175
LHX1	chr17	35294475	35294477	cg05527869	6.2955	17.2772	8.23E-14	1.78E-11	0.0363	0.7111	0.2070	0.0662	0.0118
TRABD	chr22	50623691	50623693	cg11213574	6.2140	16.0442	3.47E-13	5.17E-11	0.0385	0.7098	0.3554	0.2732	0.0196
GMDS	chr6	1625573	1625575	cg04497116	6.9296	13.3038	1.22E-11	7.96E-10	0.0218	0.7092	0.5299	0.6931	0.0308
HOXD9	chr2	176987917	176987919	cg22674699	5.6938	11.8909	9.74E-11	4.23E-09	0.0601	0.7080	0.4041	0.3027	0.0298
N/A	chr13	101194646	101194648	cg20198393	5.6770	17.6026	5.72E-14	1.37E-11	0.2312	0.7078	0.9431	0.9248	0.9199
N/A	chr7	149917262	149917264	cg02864844	7.1491	20.0208	4.54E-15	2.29E-12	0.0182	0.7065	0.1423	0.0487	0.0170
N/A	chr2	105479053	105479055	cg11014373	6.6435	15.7198	5.14E-13	6.90E-11	0.0267	0.7063	0.0925	0.0612	0.0100
TSHZ3	chr19	31841951	31841953	cg15711268	6.6028	18.1317	3.20E-14	9.00E-12	0.0275	0.7057	0.1291	0.0713	0.0139
N/A	chr3	194018157	194018159	cg16924010	5.0652	10.9612	4.24E-10	1.43E-08	0.1490	0.7053	0.8427	0.4732	0.1188
VWC2	chr7	49813101	49813103	cg02467990	6.5478	17.6419	5.48E-14	1.32E-11	0.0285	0.7046	0.4017	0.1212	0.0218
VAT1L	chr16	77822418	77822420	cg07821427	7.6074	21.1462	1.53E-15	1.08E-12	0.0128	0.7043	0.2112	0.0341	0.0117
N/A	chr13	23733861	23733863	cg05494604	6.9672	19.4595	7.96E-15	3.38E-12	0.0206	0.7040	0.2564	0.0636	0.0184
CALML3	chr10	5567477	5567479	cg20691436	6.3243	13.4253	1.03E-11	6.95E-10	0.2630	0.7032	0.9746	0.9766	0.9472
ADD2	chr2	70995458	70995460	cg15170605	6.7227	19.4657	7.91E-15	3.37E-12	0.0246	0.7025	0.2279	0.0560	0.0191
GSG1L	chr16	28074461	28074463	cg10471437	7.2613	31.0663	6.52E-19	4.48E-15	0.0164	0.7023	0.1960	0.0309	0.0135
IRF4	chr6	393238	393240	cg21277995	7.6165	34.7276	6.71E-20	8.34E-16	0.0126	0.7018	0.2845	0.0391	0.0100
FRMD4B	chr3	69591658	69591660	cg01360618	5.5942	20.4965	2.85E-15	1.69E-12	0.0629	0.7015	0.4079	0.1300	0.0452
N/A	chr17	78865086	78865088	cg10035831	5.2140	14.5906	2.14E-12	2.03E-10	0.2029	0.7014	0.8951	0.9007	0.8697
ADAMTS5	chr21	28339906	28339908	cg21646598	6.2801	13.0118	1.85E-11	1.11E-09	0.0346	0.7013	0.2916	0.1063	0.0279
N/A	chr1	9400738	9400740	cg04778236	5.7995	13.8400	5.82E-12	4.42E-10	0.2469	0.7012	0.9662	0.9599	0.9370
N/A	chr1	116371870	116371872	cg23952578	6.2252	9.7121	3.54E-09	8.51E-08	0.2631	0.7008	0.9735	0.9640	0.9476
RP11-567J20.2 lincRNA	chr8	49468827	49468829	cg18991611	7.7178	36.8880	1.95E-20	4.40E-16	0.0116	0.7005	0.2689	0.0535	0.0102

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
CTNND2	chr5	11904109	11904111	cg04996219	7.3760	22.3495	5.07E-16	4.73E-13	0.0149	0.7002	0.2414	0.0587	0.0109
NTRK3	chr15	88800566	88800568	cg05901579	6.1253	28.8734	2.89E-18	1.36E-14	0.0388	0.6991	0.2658	0.0787	0.0135
N/A	chr6	134378698	134378700	cg00583304	5.3407	13.8402	5.82E-12	4.42E-10	0.2211	0.6989	0.9262	0.9131	0.9164
N/A	chr1	246931895	246931897	cg06613253	5.3201	16.7414	1.52E-13	2.81E-11	0.2199	0.6986	0.9323	0.9191	0.9064
Top 100 hypo-methylated sites from the N v BE comparison (differentiating Barrett's esophagus from normal squamous epithelium)													
XXYLT1-AS2 antisense	chr3	194868749	194868751	cg21937377	-5.7252	-9.1804	9.23E-09	1.93E-07	0.9072	-0.7513	0.6947	0.2261	0.0840
N/A	chr13	114812779	114812781	cg08707963	-5.7178	-9.8628	2.71E-09	6.81E-08	0.9098	-0.7490	0.7258	0.8046	0.9250
N/A	chr19	2278617	2278619	cg27341866	-5.5779	-17.9023	4.11E-14	1.07E-11	0.9259	-0.7186	0.2126	0.8822	0.8772
ARID5A	chr2	97204400	97204402	cg19954537	-6.3760	-8.1279	6.85E-08	1.08E-06	0.7441	-0.7103	0.2891	0.1339	0.0235
INPP5B	chr1	38412683	38412685	cg17949727	-6.0032	-8.5898	2.79E-08	5.00E-07	0.7457	-0.7020	0.1169	0.0530	0.1055
PRR26	chr10	695858	695860	cg01263942	-5.0818	-15.1879	9.95E-13	1.14E-10	0.8894	-0.6975	0.2614	0.1967	0.8070
N/A	chr1	3086486	3086488	cg01961086	-5.7763	-9.4810	5.35E-09	1.21E-07	0.9505	-0.6911	0.4301	0.2475	0.9651
N/A	chr16	54210495	54210497	cg00253658	-4.8472	-8.7281	2.15E-08	3.98E-07	0.8427	-0.6858	0.2771	0.1451	0.3380
N/A	chr21	36577538	36577540	cg09200260	-4.7387	-13.3447	1.15E-11	7.62E-10	0.8186	-0.6740	0.5718	0.1119	0.0756
N/A	chr3	169384409	169384411	cg04120686	-4.7236	-17.7613	4.80E-14	1.20E-11	0.8055	-0.6700	0.4131	0.1603	0.0521
GDF15	chr19	18497142	18497144	cg12008047	-5.6913	-13.7097	6.96E-12	5.10E-10	0.9535	-0.6693	0.3536	0.3217	0.9402
WNK2	chr9	95948058	95948060	cg13563298	-4.7881	-13.1358	1.55E-11	9.63E-10	0.8873	-0.6655	0.3663	0.8514	0.9049
N/A	chr1	8143890	8143892	cg19274341	-4.7187	-12.4025	4.50E-11	2.26E-09	0.7792	-0.6610	0.2164	0.2048	0.3203
INPP5D	chr2	233924929	233924931	cg00438740	-4.5739	-7.8323	1.23E-07	1.80E-06	0.8176	-0.6592	0.4593	0.6815	0.0259
N/A	chr21	36577637	36577639	cg21172011	-5.0952	-15.0315	1.21E-12	1.32E-10	0.7325	-0.6583	0.2853	0.0789	0.1063
N/A	chr10	134588039	134588041	cg09893465	-5.3906	-6.7894	1.08E-06	1.19E-05	0.9453	-0.6535	0.3326	0.2836	0.9407
N/A	chr4	186434992	186434994	cg25581330	-4.9266	-13.4871	9.46E-12	6.49E-10	0.7379	-0.6532	0.1176	0.1204	0.0613
N/A	chr5	171782206	171782208	cg10505610	-5.2981	-13.8526	5.72E-12	4.36E-10	0.9412	-0.6519	0.4946	0.2640	0.9769
N/A	chr12	68848993	68848995	cg15085883	-4.5097	-12.4477	4.21E-11	2.14E-09	0.7950	-0.6495	0.4401	0.1889	0.5156
N/A	chr11	69264656	69264658	cg26540315	-4.6123	-11.1106	3.33E-10	1.17E-08	0.7691	-0.6492	0.3372	0.1772	0.1022
LAIR1	chr19	54876727	54876729	cg21878746	-4.4749	-6.2946	3.19E-06	3.06E-05	0.8396	-0.6491	0.4442	0.7485	0.0554
RNF126P1 pseudogene	chr17	55122827	55122829	cg02486253	-4.4809	-4.9806	6.44E-05	0.000424076	0.8448	-0.6488	0.2897	0.1549	0.0830
MSMB	chr10	51549504	51549506	cg10326726	-4.6691	-13.1513	1.52E-11	9.46E-10	0.8908	-0.6480	0.2547	0.2317	0.8917
LNX1	chr4	54374248	54374250	cg07080653	-5.1022	-17.6505	5.42E-14	1.32E-11	0.9320	-0.6467	0.6127	0.3824	0.9549
LAIR1	chr19	54876748	54876750	cg01515802	-4.6504	-6.1891	4.03E-06	3.76E-05	0.8897	-0.6466	0.5165	0.7837	0.0349
SSBP3	chr1	54822030	54822032	cg00172603	-4.4707	-11.2796	2.54E-10	9.36E-09	0.8606	-0.6428	0.2791	0.2337	0.0972
N/A	chr2	8849961	8849963	cg10900049	-6.3796	-22.2313	5.64E-16	5.15E-13	0.9769	-0.6404	0.3620	0.9587	0.9695
N/A	chr20	22562477	22562479	cg20504791	-4.5325	-8.9006	1.55E-08	3.01E-07	0.7611	-0.6401	0.5212	0.0951	0.1992

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
N/A	chr3	124711100	124711102	cg19533443	-4.9321	-18.5886	1.96E-14	6.35E-12	0.9237	-0.6397	0.3652	0.5152	0.8713
N/A	chr2	238578490	238578492	cg22734058	-4.5160	-10.9431	4.36E-10	1.47E-08	0.8786	-0.6383	0.3121	0.4045	0.9139
N/A	chr4	186425696	186425698	cg25999637	-5.2387	-14.4591	2.54E-12	2.32E-10	0.9432	-0.6378	0.4999	0.8264	0.9599
N/A	chr8	108208343	108208345	cg21186098	-4.5467	-11.2001	2.88E-10	1.04E-08	0.7510	-0.6367	0.1996	0.4329	0.2910
N/A	chr11	111154825	111154827	cg21127079	-5.0124	-16.1774	2.95E-13	4.59E-11	0.9324	-0.6331	0.7641	0.4996	0.9574
N/A	chr5	76034886	76034888	cg24722577	-4.4050	-13.5203	9.03E-12	6.25E-10	0.7668	-0.6324	0.0993	0.3956	0.0785
N/A	chr1	3139866	3139868	cg15727188	-4.4812	-10.1772	1.57E-09	4.28E-08	0.8821	-0.6312	0.2951	0.2725	0.8599
SPG21	chr15	65277976	65277978	cg10207553	-4.3205	-8.7903	1.91E-08	3.59E-07	0.7836	-0.6302	0.5324	0.2044	0.2336
RASA3-IT1 sense_intronic	chr13	114876209	114876211	cg21749794	-4.2860	-7.4421	2.73E-07	3.59E-06	0.7979	-0.6296	0.6154	0.6675	0.0641
MCM3AP	chr21	47704250	47704252	cg14211055	-4.6064	-12.0449	7.70E-11	3.49E-09	0.9007	-0.6294	0.6194	0.8415	0.9456
N/A	chr1	204256845	204256847	cg25407979	-4.4511	-11.9400	9.03E-11	3.98E-09	0.8802	-0.6287	0.5254	0.2149	0.1641
SERPINA11	chr14	94917626	94917628	cg21007342	-4.4493	-17.4933	6.46E-14	1.49E-11	0.7461	-0.6275	0.1383	0.1462	0.2459
N/A	chr6	112340332	112340334	cg02686793	-4.4250	-11.3301	2.34E-10	8.76E-09	0.8780	-0.6271	0.4746	0.2716	0.9311
GPSM3 / NOTCH4	chr6	32164800	32164802	cg06023661	-4.8365	-9.7050	3.58E-09	8.60E-08	0.7021	-0.6259	0.1612	0.1199	0.3821
PDE7B	chr6	136174533	136174535	cg14623715	-4.3247	-14.5404	2.29E-12	2.14E-10	0.8595	-0.6256	0.3714	0.6836	0.6437
N/A	chr3	112903053	112903055	cg00584840	-4.2726	-10.6932	6.58E-10	2.06E-08	0.8495	-0.6235	0.2711	0.1339	0.9054
N/A	chr10	115141826	115141828	cg01246622	-4.9840	-12.8858	2.22E-11	1.28E-09	0.6879	-0.6228	0.4296	0.4565	0.0668
N/A	chr4	6324628	6324630	cg16616514	-4.2692	-6.0251	5.81E-06	5.18E-05	0.8507	-0.6226	0.6921	0.7551	0.1038
N/A	chr1	111189870	111189872	cg03257417	-4.7277	-8.6207	2.63E-08	4.75E-07	0.7055	-0.6226	0.3511	0.1311	0.3610
N/A	chr10	104556331	104556333	cg09671951	-4.7575	-7.7830	1.36E-07	1.96E-06	0.6979	-0.6192	0.4162	0.1185	0.0809
INPP5D	chr2	233924922	233924924	cg02788013	-4.2367	-5.3210	2.91E-05	0.000211719	0.7692	-0.6190	0.4894	0.6917	0.0310
N/A	chr8	2037873	2037875	cg17583158	-4.3654	-12.3708	4.72E-11	2.35E-09	0.8797	-0.6179	0.4530	0.5885	0.7511
N/A	chr2	16210107	16210109	cg03687070	-4.4189	-8.3566	4.38E-08	7.35E-07	0.8880	-0.6175	0.4289	0.2703	0.9570
RP11-498C9.17 lincRNA	chr17	79924771	79924773	cg08574915	-4.1685	-15.3542	8.08E-13	9.68E-11	0.8303	-0.6164	0.2972	0.5579	0.3073
N/A	chr16	3211518	3211520	cg05982460	-5.1233	-14.3199	3.06E-12	2.67E-10	0.9447	-0.6157	0.4497	0.6811	0.9578
N/A	chr8	37218067	37218069	cg00576689	-4.6455	-12.9240	2.10E-11	1.23E-09	0.9152	-0.6139	0.3617	0.4693	0.9411
CLIC6	chr21	36041682	36041684	cg11528328	-4.1263	-5.4497	2.16E-05	0.000162929	0.8033	-0.6138	0.5940	0.5330	0.0626
SEMA3D	chr7	84671581	84671583	cg02607810	-4.1188	-16.8865	1.29E-13	2.48E-11	0.8055	-0.6130	0.5737	0.3838	0.7743
N/A	chr16	83869926	83869928	cg09683440	-4.2132	-9.6273	4.11E-09	9.67E-08	0.7564	-0.6130	0.3438	0.1502	0.0694
N/A	chr10	3160981	3160983	cg02579140	-6.2766	-16.5897	1.81E-13	3.19E-11	0.9781	-0.6129	0.4978	0.9574	0.9823
N/A	chr7	79964737	79964739	cg22110839	-4.1895	-15.3166	8.47E-13	1.00E-10	0.7619	-0.6127	0.1472	0.1527	0.5017
RP11-61F12.1	chr16	84628968	84628970	cg03738384	-4.3636	-11.4709	1.87E-10	7.28E-09	0.8852	-0.6127	0.2475	0.2505	0.8560

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
antisense													
N/A	chr6	39176488	39176490	cg21619775	-4.6300	-12.9628	1.98E-11	1.17E-09	0.9146	-0.6126	0.4134	0.4665	0.9043
N/A	chr11	268949	268951	cg13185005	-4.3436	-16.0007	3.65E-13	5.36E-11	0.7279	-0.6115	0.2285	0.4909	0.0524
PIGB	chr15	55610623	55610625	cg19991916	-4.5713	-11.1186	3.28E-10	1.16E-08	0.9107	-0.6104	0.7502	0.8486	0.9355
N/A	chr13	114880888	114880890	cg15972148	-4.2766	-13.5477	8.70E-12	6.07E-10	0.8762	-0.6087	0.4216	0.6814	0.5137
N/A	chr10	3160925	3160927	cg05394294	-4.7193	-19.3696	8.73E-15	3.60E-12	0.9241	-0.6081	0.4321	0.8718	0.9194
RP11-443B7.2 lincRNA	chr1	235091105	235091107	cg20155035	-4.0721	-17.9342	3.97E-14	1.05E-11	0.7972	-0.6078	0.1765	0.3989	0.2108
IRF8	chr16	85933162	85933164	cg04597312	-4.1190	-14.8322	1.57E-12	1.60E-10	0.7646	-0.6071	0.1703	0.1628	0.7041
N/A	chr5	78277491	78277493	cg08393937	-4.0503	-16.9875	1.14E-13	2.26E-11	0.7969	-0.6054	0.2924	0.2444	0.7274
N/A	chr15	78726575	78726577	cg20117256	-4.1149	-12.6873	2.96E-11	1.61E-09	0.7573	-0.6047	0.2829	0.4099	0.2888
N/A	chr16	81328088	81328090	cg00602295	-4.0825	-10.4646	9.65E-10	2.84E-08	0.7645	-0.6037	0.1966	0.1858	0.3247
LINC00261 lincRNA	chr20	22557517	22557519	cg08582485	-4.0612	-8.8072	1.85E-08	3.50E-07	0.7734	-0.6037	0.4014	0.1164	0.1979
SLC6A20	chr3	45837196	45837198	cg24940967	-4.0671	-9.3141	7.23E-09	1.57E-07	0.8333	-0.6036	0.4712	0.5924	0.4450
FLG2	chr1	152333422	152333424	cg03957898	-4.0406	-11.7463	1.22E-10	5.09E-09	0.8244	-0.6024	0.4717	0.3626	0.8028
N/A	chr3	169384033	169384035	cg09380805	-4.0200	-9.7771	3.15E-09	7.73E-08	0.7870	-0.6015	0.4798	0.1825	0.0898
N/A	chr1	12603853	12603855	cg20059492	-4.5082	-11.5663	1.61E-10	6.42E-09	0.9113	-0.6002	0.4353	0.3485	0.9403
CLIC6	chr21	36041604	36041606	cg19200589	-3.9998	-7.5517	2.18E-07	2.95E-06	0.8065	-0.5998	0.5746	0.4727	0.0294
N/A	chr10	123810031	123810033	cg05655953	-5.7869	-21.5344	1.07E-15	8.40E-13	0.9702	-0.5993	0.3969	0.4357	0.9451
INMT	chr7	30737555	30737557	cg13134297	-4.0843	-13.6018	8.07E-12	5.73E-10	0.7485	-0.5992	0.2661	0.4839	0.2834
IDO1	chr8	39782017	39782019	cg24188163	-4.0122	-17.9939	3.72E-14	1.00E-11	0.7751	-0.5991	0.3325	0.1948	0.7816
FLNB	chr3	58151254	58151256	cg05116002	-4.3762	-9.9915	2.17E-09	5.62E-08	0.8992	-0.5987	0.6474	0.2883	0.9267
N/A	chr2	239403002	239403004	cg03217587	-4.0047	-12.7428	2.73E-11	1.51E-09	0.8216	-0.5987	0.2144	0.6836	0.0663
PRR26	chr10	695876	695878	cg17198320	-4.2225	-16.3143	2.51E-13	4.07E-11	0.8798	-0.5982	0.4039	0.2511	0.7990
N/A	chr3	169377724	169377726	cg08004073	-3.9818	-9.7736	3.17E-09	7.77E-08	0.7991	-0.5980	0.3229	0.2549	0.0760
N/A	chr6	122253934	122253936	cg15010554	-4.3981	-23.1047	2.61E-16	3.00E-13	0.6936	-0.5967	0.2625	0.1152	0.1863
N/A	chr8	122823861	122823863	cg10131026	-3.9745	-9.5382	4.82E-09	1.11E-07	0.8106	-0.5966	0.6546	0.4002	0.5814
N/A	chr14	70160159	70160161	cg15935770	-3.9882	-11.7800	1.16E-10	4.88E-09	0.7734	-0.5964	0.2505	0.4254	0.3750
N/A	chr7	155592262	155592264	cg00802903	-3.9842	-4.7091	0.000122277	0.00074227	0.8257	-0.5953	0.7369	0.7637	0.6551
N/A	chr15	61044828	61044830	cg02010763	-4.1813	-9.6414	4.01E-09	9.47E-08	0.8768	-0.5950	0.5388	0.2084	0.8698
N/A	chr11	85453668	85453670	cg20943999	-4.8256	-9.1900	9.07E-09	1.90E-07	0.9365	-0.5942	0.4244	0.3910	0.9285
N/A	chr2	127955155	127955157	cg24923516	-4.0708	-13.3594	1.13E-11	7.49E-10	0.8578	-0.5936	0.4572	0.8108	0.3561
RP11-432J24.2 lincRNA	chr10	134231548	134231550	cg23249922	-4.1361	-12.3925	4.57E-11	2.29E-09	0.8710	-0.5936	0.5497	0.3475	0.4207

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
N/A	chr19	1547395	1547397	cg05649427	-4.4526	-14.6824	1.90E-12	1.86E-10	0.9107	-0.5929	0.5714	0.6729	0.9368
PRR26	chr10	696331	696333	cg14121234	-6.3214	-17.2241	8.74E-14	1.86E-11	0.9806	-0.5928	0.4200	0.3788	0.9881
ST14	chr11	130059790	130059792	cg23250795	-4.2267	-14.1700	3.73E-12	3.12E-10	0.8859	-0.5928	0.3364	0.6018	0.8884
N/A	chr17	70120540	70120542	cg06234051	-5.4445	-12.9736	1.95E-11	1.16E-09	0.9624	-0.5922	0.4268	0.3458	0.9293
RP11-797H7.5 antisense	chr7	64349132	64349134	cg13990351	-4.1741	-7.0007	6.89E-07	8.03E-06	0.7123	-0.5917	0.0393	0.0239	0.0176
N/A	chr4	124521390	124521392	cg03862787	-4.4524	-18.8392	1.51E-14	5.23E-12	0.6778	-0.5901	0.1423	0.1150	0.0670
TNS3	chr7	47493387	47493389	cg07241084	-3.9397	-7.2895	3.75E-07	4.73E-06	0.7616	-0.5893	0.3066	0.1682	0.1261
C1orf132 processed_transcript	chr1	207997019	207997021	cg10501210	-3.9050	-11.5966	1.54E-10	6.18E-09	0.7892	-0.5893	0.2792	0.2417	0.8449
ARHGAP44	chr17	12695724	12695726	cg24096540	-5.1677	-15.6439	5.64E-13	7.36E-11	0.9537	-0.5892	0.4151	0.7847	0.9657

Table 3: Top 100 hyper- and hypo-methylated probes for differentiating intervention requiring disease from Barrett's esophagus
1,648 hypermethylated and 153 hypomethylated CpG sites were identified as differentially methylated between Barrett's esophagus and intervention requiring disease (BE v HGD-EAC, FDR < 0.01 and Δβ ≥ 0.20). Gencode v19 was used for annotation[204], gene name protein-coding when not specified. logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, β_{base}: average methylation in baseline samples, here, Barrett's esophagus, Δβ: difference in average methylation between comparison groups, here, between BE and HGD-EAC, β_{duo}: average methylation in duodenal epithelium, β_{stom}: average methylation in proximal stomach epithelium, β_{blood}: average methylation in peripheral blood from disease-free patients. Top 100 hyper- and hypo-methylated sites based on greatest |Δβ|; listed in order of decreasing |Δβ|.

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
Top 100 hyper-methylated sites from the BE v HGD-EAC comparison (differentiating intervention requiring disease from Barrett's esophagus)													
N/A	chr13	114812779	114812781	cg08707963	4.0335	6.3513	2.81E-06	0.002321031	0.1609	0.5975	0.7258	0.8046	0.9250
DTX3	chr12	57998785	57998787	cg17730484	4.0435	7.4563	2.65E-07	0.000654129	0.1189	0.5710	0.0488	0.0563	0.0465
DCBLD2	chr3	98620661	98620663	cg27352765	5.4414	6.6768	1.38E-06	0.001614185	0.0318	0.5564	0.0144	0.0168	0.0194
DCHS1	chr11	6677086	6677088	cg21012362	4.6033	7.3346	3.41E-07	0.000789165	0.0621	0.5546	0.0133	0.0227	0.0192
SNORD103B snoRNA	chr1	31423070	31423072	cg04248332	3.8322	6.5539	1.80E-06	0.001813045	0.3119	0.5540	0.7152	0.4085	0.9113
GBGT1	chr9	136039124	136039126	cg14472025	5.0940	11.5566	1.64E-10	1.25E-05	0.0392	0.5428	0.0246	0.0384	0.0237
DTX3	chr12	57998761	57998763	cg11654179	3.6754	7.6312	1.85E-07	0.000565773	0.1380	0.5336	0.0527	0.0691	0.0203
ENOX1	chr13	44361211	44361213	cg24098938	3.7498	6.1753	4.15E-06	0.00281916	0.1223	0.5298	0.0191	0.0117	0.0127
GNG4	chr1	235812839	235812841	cg06714284	4.2635	7.5149	2.35E-07	0.000618223	0.0727	0.5283	0.0100	0.0123	0.0100
N/A	chr5	36690258	36690260	cg08407114	3.5717	7.5789	2.06E-07	0.000582086	0.1420	0.5211	0.2623	0.1483	0.0296

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
ENOX1	chr13	44361219	44361221	cg01118451	3.5284	5.6312	1.42E-05	0.00501497	0.1489	0.5199	0.0283	0.0272	0.0102
N/A	chr5	36690456	36690458	cg18064852	3.5381	9.3310	7.01E-09	8.66E-05	0.1377	0.5120	0.1742	0.0781	0.0567
N/A	chr5	36690656	36690658	cg00132108	3.9272	10.8398	5.17E-10	2.00E-05	0.0890	0.5088	0.1063	0.0931	0.0584
FAM78B	chr1	166134281	166134283	cg21868774	3.5507	6.0964	4.95E-06	0.003050374	0.1302	0.5067	0.0172	0.0262	0.0128
EMILIN2	chr18	2847529	2847531	cg05102581	3.8930	6.0816	5.12E-06	0.003087359	0.0907	0.5064	0.1153	0.0171	0.0103
N/A	chr5	36690436	36690438	cg05542957	5.6720	13.8895	5.44E-12	2.50E-06	0.0210	0.5014	0.0564	0.0138	0.0155
KCNB2	chr8	73449523	73449525	cg15774465	4.0321	6.1592	4.31E-06	0.00287227	0.0757	0.4969	0.0621	0.0274	0.0108
QPCT	chr2	37571728	37571730	cg15162392	4.3648	5.2101	3.77E-05	0.008126585	0.0565	0.4960	0.0465	0.0637	0.0373
ADCY5	chr3	123167769	123167771	cg02978184	3.3560	5.2627	3.33E-05	0.007655235	0.1523	0.4956	0.0367	0.0305	0.0125
RARA	chr17	38506349	38506351	cg18125573	3.4858	5.5506	1.71E-05	0.005558375	0.3758	0.4951	0.8638	0.3560	0.9426
CLDN10	chr13	96204872	96204874	cg10305311	3.8961	6.7417	1.20E-06	0.001510758	0.0844	0.4941	0.1449	0.0236	0.0114
AC009236.2 lincRNA	chr2	45395869	45395871	cg20459525	3.3935	8.0388	8.17E-08	0.000351035	0.1421	0.4930	0.0450	0.0932	0.0634
RRAGD	chr6	90121669	90121671	cg22264975	3.5273	10.2153	1.47E-09	3.23E-05	0.1204	0.4917	0.0447	0.0780	0.0304
NMUR1	chr2	232395268	232395270	cg19187155	3.4086	5.3112	2.98E-05	0.007290931	0.1367	0.4903	0.0668	0.0704	0.0208
ENO4	chr10	118609166	118609168	cg15461105	4.7516	11.3811	2.16E-10	1.42E-05	0.0397	0.4870	0.0233	0.0445	0.0259
KCNQ3	chr8	133493433	133493435	cg04396550	3.2384	5.5880	1.57E-05	0.005319329	0.1653	0.4861	0.0191	0.0219	0.0139
ZBTB20	chr3	114103344	114103346	cg13792714	3.1260	5.4940	1.95E-05	0.00590825	0.1969	0.4846	0.6993	0.5548	0.7404
OBSL1	chr2	220417248	220417250	cg11531021	3.1584	5.3670	2.61E-05	0.006897738	0.1846	0.4844	0.0239	0.0271	0.0133
ENO4	chr10	118608697	118608699	cg01035689	3.5355	9.7828	3.12E-09	5.32E-05	0.1105	0.4798	0.0698	0.1111	0.0834
KCNMA1	chr10	79397454	79397456	cg09405661	3.0217	5.1978	3.87E-05	0.008231347	0.2387	0.4793	0.0388	0.0468	0.0313
GRASP	chr12	52400662	52400664	cg22196952	5.1079	9.1412	9.92E-09	0.000108173	0.0291	0.4788	0.0124	0.0121	0.0100
GPR25	chr1	200842281	200842283	cg19940077	3.1130	8.2719	5.16E-08	0.000264917	0.3380	0.4774	0.6155	0.2963	0.8460
N/A	chr11	130641100	130641102	cg16864819	3.2846	7.8211	1.26E-07	0.000456997	0.3808	0.4762	0.7347	0.6764	0.8865
PODN	chr1	53527664	53527666	cg01394819	2.9963	5.5760	1.61E-05	0.005401726	0.2261	0.4737	0.0316	0.0205	0.0133
KCNB2	chr8	73449518	73449520	cg18555069	3.8960	6.6330	1.52E-06	0.001696913	0.0754	0.4728	0.0765	0.0113	0.0102
USP32	chr17	58498976	58498978	cg09695735	4.8985	5.1528	4.30E-05	0.008596317	0.0327	0.4691	0.0409	0.0100	0.0100
QPCT	chr2	37571731	37571733	cg08786077	3.6360	5.6657	1.31E-05	0.004824456	0.0938	0.4690	0.0873	0.1146	0.0563
SH3RF3	chr2	109745613	109745615	cg22959667	3.9233	6.4746	2.14E-06	0.001987147	0.0716	0.4677	0.0124	0.0193	0.0103
INSC	chr11	15136455	15136457	cg02434443	3.3785	7.5072	2.39E-07	0.00062276	0.1203	0.4668	0.1685	0.0340	0.0131
SH3RF3	chr2	109745647	109745649	cg16117910	3.6117	5.4943	1.95E-05	0.00590825	0.0947	0.4664	0.0211	0.0380	0.0162
N/A	chr17	76993393	76993395	cg00902147	3.3416	7.4562	2.65E-07	0.000654129	0.4100	0.4657	0.9205	0.7199	0.9834
PRKG1	chr10	52833908	52833910	cg04884011	3.1050	5.4457	2.18E-05	0.006203342	0.1659	0.4653	0.0803	0.0332	0.0275
N/A	chr1	247611501	247611503	cg09226051	3.3354	5.8112	9.43E-06	0.00409674	0.1218	0.4615	0.1542	0.0717	0.2766
SCARF2	chr22	20792216	20792218	cg15243570	3.1402	7.6303	1.86E-07	0.000565773	0.1532	0.4615	0.0386	0.0322	0.0100

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
PCLO	chr7	82792344	82792346	cg15478390	3.7534	7.2313	4.24E-07	0.000892687	0.0805	0.4610	0.0307	0.0144	0.0194
WWP2	chr16	69924615	69924617	cg27456203	2.8969	6.2708	3.36E-06	0.00248741	0.2307	0.4600	0.5828	0.1931	0.8501
CTD-3194G12.2 lincRNA	chr17	35078658	35078660	cg21653132	3.4578	5.8886	7.91E-06	0.003761777	0.1062	0.4600	0.0352	0.0281	0.0178
N/A	chr19	36478288	36478290	cg13620744	3.1270	5.0557	5.40E-05	0.009586952	0.1508	0.4573	0.0302	0.0239	0.0144
DNM3	chr1	171810777	171810779	cg06211893	3.0632	5.9481	6.92E-06	0.003578714	0.1632	0.4566	0.1392	0.0372	0.0169
KRBA1	chr7	149412289	149412291	cg03788131	3.4640	5.5918	1.56E-05	0.005286613	0.1029	0.4557	0.1176	0.0322	0.0270
FADS1	chr11	61583570	61583572	cg13475388	4.2667	5.5058	1.90E-05	0.005854958	0.0505	0.4552	0.0136	0.0111	0.0119
N/A	chr17	31149876	31149878	cg01050010	3.0708	5.2913	3.12E-05	0.007422554	0.1590	0.4547	0.3245	0.3553	0.2474
FADS1	chr11	61583955	61583957	cg23992449	3.8550	5.8032	9.60E-06	0.004119001	0.0710	0.4541	0.0109	0.0143	0.0103
N/A	chr19	11784760	11784762	cg21771200	3.4896	6.6041	1.62E-06	0.001743736	0.0993	0.4540	0.4908	0.0540	0.0555
STOX2	chr4	184826298	184826300	cg10957242	3.3215	9.1851	9.15E-09	0.000107932	0.1152	0.4503	0.0227	0.0633	0.0187
EPHX3	chr19	15343344	15343346	cg17399362	3.5499	6.4958	2.05E-06	0.001926668	0.0916	0.4500	0.0725	0.0786	0.0721
ATP6V1B1	chr2	71192261	71192263	cg10598816	2.8355	5.3105	2.98E-05	0.007291554	0.2260	0.4498	0.0538	0.0642	0.0510
ENOX1	chr13	44361018	44361020	cg09320746	3.0738	6.2817	3.28E-06	0.002459659	0.1522	0.4496	0.0533	0.0538	0.0513
N/A	chr22	39784480	39784482	cg02038168	3.2013	5.0292	5.75E-05	0.009898916	0.1307	0.4496	0.0921	0.0340	0.0264
TTYH1	chr19	54926616	54926618	cg15723536	2.9234	5.5630	1.66E-05	0.005462332	0.1876	0.4490	0.1494	0.0631	0.0193
MOXD1	chr6	132722420	132722422	cg07570142	3.5723	9.8867	2.60E-09	4.99E-05	0.0891	0.4487	0.1076	0.0690	0.0687
RET	chr10	43572243	43572245	cg00540891	2.8762	5.8240	9.16E-06	0.004023324	0.2008	0.4477	0.0518	0.0922	0.0293
N/A	chr5	64103335	64103337	cg15663823	2.8199	6.5234	1.93E-06	0.00188473	0.3319	0.4463	0.7027	0.3887	0.9056
MEIS1	chr2	66667548	66667550	cg11433622	3.0141	5.3458	2.75E-05	0.00701948	0.1600	0.4461	0.2529	0.3408	0.0282
HLX	chr1	221053512	221053514	cg15356516	3.8118	5.9238	7.30E-06	0.003639877	0.0703	0.4448	0.0344	0.0512	0.0214
N/A	chr1	76081961	76081963	cg27547954	2.7569	5.5458	1.73E-05	0.005587535	0.2756	0.4444	0.0534	0.0100	0.0111
TTC9B	chr19	40724360	40724362	cg02724271	3.1309	5.6329	1.42E-05	0.005013328	0.1363	0.4440	0.0340	0.0214	0.0145
SCN5A	chr3	38691363	38691365	cg12926589	2.9502	6.3781	2.65E-06	0.002246071	0.1704	0.4432	0.0630	0.0825	0.0895
MOXD1	chr6	132722314	132722316	cg16478774	4.9743	11.6106	1.50E-10	1.25E-05	0.0275	0.4431	0.0446	0.0119	0.0100
PCP4L1	chr1	161228495	161228497	cg27223727	3.7980	6.0661	5.30E-06	0.003118637	0.0702	0.4421	0.0139	0.0180	0.0119
FADS1	chr11	61584111	61584113	cg27173322	3.6321	5.0862	5.03E-05	0.009246452	0.0809	0.4408	0.0150	0.0100	0.0100
EPHX3	chr19	15343394	15343396	cg19744936	3.0748	5.7771	1.02E-05	0.004231267	0.1421	0.4405	0.1426	0.1051	0.0752
DNM3	chr1	171810569	171810571	cg14376275	2.9018	7.0967	5.62E-07	0.001009502	0.1777	0.4399	0.1308	0.0703	0.0244
GRASP	chr12	52400666	52400668	cg09101796	4.5903	7.4387	2.75E-07	0.000669632	0.0362	0.4391	0.0161	0.0117	0.0100
RTN1	chr14	60337476	60337478	cg19556814	4.3616	5.2812	3.19E-05	0.007501944	0.0430	0.4372	0.0218	0.0362	0.0107
SCARF2	chr22	20792221	20792223	cg03952331	3.0685	7.9038	1.07E-07	0.000430591	0.1383	0.4355	0.0158	0.0145	0.0100
BOC	chr3	112930674	112930676	cg27085904	4.1660	5.6038	1.51E-05	0.005211485	0.0499	0.4353	0.0550	0.0544	0.0339

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
N/A	chr7	127807499	127807501	cg25402610	3.4832	5.3360	2.81E-05	0.00710729	0.0897	0.4346	0.0105	0.0176	0.0100
N/A	chr12	121154736	121154738	cg23366404	3.0373	8.1111	7.08E-08	0.000320913	0.4240	0.4340	0.8252	0.9165	0.9306
MIR23B miRNA	chr9	97846550	97846552	cg00351472	2.7375	7.7354	1.50E-07	0.000499938	0.3430	0.4339	0.8012	0.7944	0.8788
N/A	chr20	17595447	17595449	cg26161708	2.6825	6.6099	1.60E-06	0.001736403	0.2720	0.4338	0.4220	0.2588	0.9597
N/A	chr10	102587109	102587111	cg26439963	3.9336	12.6939	2.93E-11	6.73E-06	0.0600	0.4337	0.0349	0.0540	0.0291
DNM3	chr1	171811476	171811478	cg20800956	3.7989	6.0195	5.89E-06	0.00329238	0.0672	0.4333	0.0618	0.0197	0.0102
N/A	chr13	80233182	80233184	cg20271865	2.7059	9.3699	6.53E-09	8.66E-05	0.2403	0.4333	0.2902	0.5132	0.8539
COL19A1	chr6	70577477	70577479	cg17840061	2.8147	5.6117	1.49E-05	0.005165124	0.1885	0.4319	0.0341	0.0401	0.0709
SLC16A14	chr2	230933066	230933068	cg26338195	4.0962	6.2716	3.35E-06	0.00248741	0.0519	0.4317	0.0266	0.0348	0.0196
ZNF365	chr10	64133895	64133897	cg08984023	3.7248	10.2745	1.33E-09	3.06E-05	0.0710	0.4315	0.0815	0.0489	0.0456
DTX3	chr12	57998655	57998657	cg19217692	3.7273	10.8362	5.20E-10	2.00E-05	0.0706	0.4309	0.0631	0.0501	0.0548
SNORD103B snoRNA	chr1	31423282	31423284	cg00581541	2.8275	6.4359	2.33E-06	0.002089962	0.3869	0.4306	0.7399	0.5569	0.9039
GRIN2A	chr16	10276798	10276800	cg16368442	3.9156	5.1558	4.27E-05	0.008564896	0.0580	0.4237	0.1323	0.0164	0.0113
GSC	chr14	95236122	95236124	cg10042799	2.7612	6.4693	2.17E-06	0.001994471	0.1893	0.4235	0.0435	0.0279	0.0100
N/A	chr11	16170022	16170024	cg21992400	2.6436	10.6847	6.68E-10	2.36E-05	0.3400	0.4230	0.5923	0.7226	0.9092
DNM3	chr1	171810375	171810377	cg04399751	2.6045	5.3669	2.61E-05	0.006897738	0.2578	0.4209	0.3779	0.1006	0.0593
DCBLD2	chr3	98620815	98620817	cg17198587	3.5096	6.3281	2.96E-06	0.002353487	0.0814	0.4208	0.0489	0.0523	0.0447
PIANP	chr12	6809511	6809513	cg12376068	2.6622	5.3429	2.76E-05	0.007044024	0.2186	0.4205	0.0452	0.0693	0.0409
DCBLD2	chr3	98620856	98620858	cg02464093	4.3855	5.1737	4.10E-05	0.008377138	0.0391	0.4205	0.0103	0.0207	0.0118
SEPW1	chr19	48285321	48285323	cg07953015	2.5903	6.9129	8.30E-07	0.001251874	0.2705	0.4202	0.3720	0.2460	0.9075
N/A	chr12	26963488	26963490	cg26381514	2.6350	6.6906	1.34E-06	0.001599275	0.3505	0.4197	0.5625	0.3171	0.9099
TRIM17	chr1	228604412	228604414	cg01907584	2.7675	5.4851	1.99E-05	0.005921511	0.1806	0.4195	0.0906	0.0921	0.0474
PIGR	chr1	207120021	207120023	cg20953047	2.8184	10.4408	1.00E-09	2.89E-05	0.4144	0.4187	0.4803	0.8615	0.9021
Top 100 hypo-methylated sites from the BE v HGD-EAC comparison (differentiating intervention requiring disease from Barrett's esophagus)													
SOCS2	chr12	93967579	93967581	cg07687131	-2.9815	-5.2577	3.37E-05	0.007687167	0.6585	-0.4623	0.2033	0.7138	0.1557
EFCAB4A	chr11	830526	830528	cg06085985	-3.0039	-5.5809	1.60E-05	0.005365279	0.5985	-0.4418	0.3232	0.3162	0.0976
RP13-452N2.1 processed_transcript	chr7	151505115	151505117	cg22528270	-2.8620	-5.4987	1.93E-05	0.005878605	0.8559	-0.4062	0.8710	0.9525	0.3927
N/A	chr8	10192942	10192944	cg06620390	-3.6395	-5.8878	7.93E-06	0.003761777	0.9396	-0.3844	0.9419	0.8852	0.6912
ZNF606	chr19	58513647	58513649	cg00664634	-2.8363	-5.8324	8.98E-06	0.00398856	0.5072	-0.3813	0.1292	0.0727	0.0501
FBLN2	chr3	13679573	13679575	cg17378193	-2.2493	-5.0307	5.73E-05	0.009882997	0.7108	-0.3700	0.6037	0.7389	0.2557
BST2	chr19	17516328	17516330	cg16363586	-2.2076	-5.1836	4.01E-05	0.00832743	0.6629	-0.3643	0.6553	0.7272	0.1780
N/A	chr6	33096302	33096304	cg15019001	-2.1914	-6.1093	4.81E-06	0.003012402	0.6906	-0.3623	0.6531	0.7279	0.5065
MMRN2	chr10	88730598	88730600	cg06744119	-2.3580	-5.5146	1.86E-05	0.005803324	0.5525	-0.3584	0.2663	0.1345	0.0774

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
N/A	chr7	136575359	136575361	cg10276834	-2.0960	-6.0366	5.67E-06	0.003221268	0.6445	-0.3467	0.6500	0.6413	0.2535
N/A	chr21	43183465	43183467	cg03079733	-2.0742	-5.9823	6.40E-06	0.003440231	0.6756	-0.3447	0.5998	0.4209	0.2149
N/A	chr3	46139304	46139306	cg02352281	-2.0508	-5.1484	4.35E-05	0.008620947	0.6759	-0.3411	0.5917	0.6851	0.2087
N/A	chr6	167763757	167763759	cg10921277	-2.7558	-6.0132	5.97E-06	0.003298436	0.4480	-0.3407	0.2345	0.3670	0.2004
N/A	chr16	7354040	7354042	cg27199384	-2.0786	-5.6095	1.49E-05	0.005183593	0.6152	-0.3406	0.1499	0.0435	0.0135
N/A	chr8	10208256	10208258	cg26966828	-2.4205	-5.3256	2.88E-05	0.0071794	0.8456	-0.3399	0.7218	0.8193	0.1752
TRGJP1	chr7	38315970	38315972	cg04978343	-1.9475	-5.3354	2.81E-05	0.007110096	0.7269	-0.3186	0.7779	0.7195	0.1297
CDKN2A	chr9	21996206	21996208	cg14069088	-1.9145	-6.5382	1.86E-06	0.001837349	0.6217	-0.3181	0.5965	0.5697	0.1930
N/A	chr10	24160304	24160306	cg03844372	-2.0402	-5.8474	8.69E-06	0.003919528	0.5233	-0.3126	0.6063	0.6534	0.0569
RP11-968A15.2 antisense	chr12	54655893	54655895	cg08269188	-1.9130	-6.7434	1.19E-06	0.001509272	0.7308	-0.3119	0.6844	0.6958	0.2191
DOK2	chr8	21771058	21771060	cg03732056	-1.9379	-6.5648	1.76E-06	0.001790566	0.5458	-0.3070	0.5060	0.4574	0.0625
N/A	chr4	109994038	109994040	cg08760493	-1.7821	-5.5404	1.75E-05	0.005617727	0.6378	-0.2992	0.7423	0.7365	0.3963
EFCAB4A	chr11	831169	831171	cg02261765	-1.7738	-5.6823	1.26E-05	0.004713968	0.6315	-0.2977	0.4122	0.4257	0.2288
RALBP1	chr18	9474142	9474144	cg21923525	-1.8020	-7.5537	2.17E-07	0.000594626	0.7162	-0.2963	0.6979	0.7455	0.3495
N/A	chr2	208014176	208014178	cg19361800	-1.8214	-5.4307	2.26E-05	0.006352666	0.7366	-0.2949	0.7676	0.6681	0.4846
UNKL	chr16	1428705	1428707	cg08063051	-1.7670	-5.3114	2.97E-05	0.007290931	0.6891	-0.2947	0.6875	0.7092	0.7376
N/A	chr4	24975753	24975755	cg17173896	-1.7638	-5.1841	4.00E-05	0.008320964	0.6887	-0.2942	0.7476	0.6892	0.4569
N/A	chr6	33096311	33096313	cg08506353	-1.9134	-5.2331	3.57E-05	0.007902748	0.5025	-0.2910	0.5638	0.6257	0.3820
N/A	chr5	42720520	42720522	cg18304305	-1.9234	-5.5939	1.55E-05	0.00527294	0.7941	-0.2899	0.7305	0.7650	0.1234
N/A	chr18	77586146	77586148	cg20450689	-1.8025	-6.5826	1.69E-06	0.001774645	0.7522	-0.2869	0.7412	0.8021	0.3760
ATP11A	chr13	113343375	113343377	cg03992114	-1.7123	-5.6844	1.26E-05	0.004703423	0.6964	-0.2846	0.6347	0.6519	0.1961
N/A	chr7	149562622	149562624	cg01077185	-1.6712	-7.0179	6.64E-07	0.001106374	0.6346	-0.2817	0.6482	0.6437	0.1951
N/A	chr16	6094219	6094221	cg08390979	-1.6915	-5.6660	1.31E-05	0.004824456	0.7001	-0.2806	0.7431	0.6695	0.8026
CEBPE	chr14	23589418	23589420	cg15691199	-1.7098	-6.9735	7.30E-07	0.001157145	0.7247	-0.2788	0.6377	0.7140	0.1033
COMMD3	chr10	22608857	22608859	cg10763374	-1.6510	-6.8516	9.47E-07	0.001335497	0.6182	-0.2780	0.5913	0.7222	0.3231
N/A	chr21	46388161	46388163	cg11327657	-3.2905	-5.2167	3.71E-05	0.008047264	0.9540	-0.2746	0.9430	0.9623	0.3733
N/A	chr12	124434451	124434453	cg02658319	-1.8462	-5.2218	3.66E-05	0.007994971	0.7997	-0.2735	0.7566	0.8013	0.8503
GPR85	chr7	112724673	112724675	cg14511782	-1.6167	-6.9915	7.02E-07	0.001133164	0.6334	-0.2730	0.6358	0.6334	0.3905
ZNF480	chr19	52801050	52801052	cg19866478	-2.2008	-7.3244	3.49E-07	0.000800043	0.3989	-0.2727	0.2536	0.1717	0.0957
IFLTD1	chr12	25707568	25707570	cg24850296	-1.6128	-5.7389	1.11E-05	0.004395291	0.6202	-0.2721	0.6172	0.6233	0.3536
ZNF239	chr10	44068713	44068715	cg07541020	-1.7259	-5.5121	1.87E-05	0.005820641	0.7567	-0.2721	0.7974	0.7932	0.4747
N/A	chr16	50743026	50743028	cg04172533	-1.7023	-5.1634	4.20E-05	0.008473117	0.7503	-0.2702	0.6713	0.7331	0.3035
N/A	chr21	46325852	46325854	cg04217515	-2.1655	-7.8767	1.13E-07	0.000432454	0.8712	-0.2699	0.8293	0.9046	0.5474

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
N/A	chr21	27942802	27942804	cg14843731	-1.5833	-6.3025	3.13E-06	0.002395345	0.6338	-0.2677	0.5774	0.4675	0.1825
ICAM3	chr19	10450021	10450023	cg14145194	-1.6572	-6.1705	4.20E-06	0.002838124	0.7336	-0.2675	0.6670	0.7325	0.1831
N/A	chr1	42217076	42217078	cg23114964	-1.5811	-5.1711	4.12E-05	0.00841023	0.6667	-0.2660	0.6012	0.6452	0.0281
ACSF2	chr17	48502311	48502313	cg03919657	-1.6214	-6.5919	1.66E-06	0.001769908	0.7150	-0.2658	0.5406	0.7818	0.4094
N/A	chr15	80765903	80765905	cg27332878	-1.6215	-8.9275	1.47E-08	0.000138427	0.7246	-0.2637	0.6451	0.6889	0.2540
TBC1D13	chr9	131549074	131549076	cg27625732	-1.7524	-6.3823	2.63E-06	0.002236317	0.7881	-0.2634	0.6811	0.6735	0.5530
N/A	chr16	18039446	18039448	cg04749323	-1.6611	-5.4730	2.05E-05	0.006004868	0.7529	-0.2622	0.8031	0.7588	0.8666
S100B	chr21	48026042	48026044	cg00025591	-1.5509	-5.3601	2.66E-05	0.006948712	0.6045	-0.2617	0.4809	0.5174	0.0771
HERC2	chr15	28448837	28448839	cg07726456	-1.7954	-5.8478	8.68E-06	0.003919528	0.8077	-0.2602	0.7490	0.8200	0.1755
AP006216.12 antisense	chr11	116706152	116706154	cg03044513	-1.6094	-5.1906	3.94E-05	0.008274415	0.7326	-0.2595	0.6561	0.7199	0.3136
AP006216.12 antisense	chr11	116706050	116706052	cg19299755	-1.9385	-6.2576	3.46E-06	0.002541402	0.8451	-0.2577	0.7888	0.8147	0.4694
RBFOX1	chr16	6533699	6533701	cg12310850	-1.5468	-5.6050	1.51E-05	0.00521066	0.5617	-0.2568	0.1969	0.1746	0.1262
RP11-109A6.3 lincRNA	chr10	131568020	131568022	cg14308082	-1.8815	-5.4941	1.95E-05	0.00590825	0.8345	-0.2567	0.8218	0.8393	0.4829
N/A	chr10	131412578	131412580	cg19680672	-1.6851	-5.7270	1.14E-05	0.004452487	0.7799	-0.2556	0.8167	0.8127	0.8731
N/A	chr3	55693630	55693632	cg18243357	-1.5252	-6.5379	1.87E-06	0.001837349	0.6988	-0.2525	0.5929	0.6949	0.6861
N/A	chr6	33142005	33142007	cg15818109	-1.6568	-5.9454	6.96E-06	0.003578714	0.7819	-0.2499	0.7583	0.7658	0.4720
MSX1	chr4	4860589	4860591	cg22371591	-1.6495	-5.2851	3.16E-05	0.007464683	0.7831	-0.2481	0.3349	0.4055	0.3070
ZNF414	chr19	8580333	8580335	cg14574996	-1.8293	-5.2513	3.42E-05	0.007731177	0.8364	-0.2465	0.7853	0.7875	0.2134
N/A	chr13	107145932	107145934	cg24128998	-1.5129	-5.8502	8.63E-06	0.003910354	0.7186	-0.2464	0.6109	0.7596	0.0775
N/A	chr5	178776242	178776244	cg11303630	-1.4621	-5.7684	1.04E-05	0.004268686	0.5798	-0.2461	0.6180	0.6244	0.2730
N/A	chr5	92909433	92909435	cg15143788	-1.4893	-5.0305	5.73E-05	0.009882997	0.7065	-0.2449	0.4218	0.4242	0.1078
P2RY12	chr3	151057060	151057062	cg24630764	-1.4907	-6.1045	4.87E-06	0.003028742	0.5337	-0.2443	0.5964	0.4683	0.5827
N/A	chr1	216805706	216805708	cg13242895	-1.6743	-5.3321	2.83E-05	0.007140762	0.8016	-0.2429	0.8370	0.8101	0.8497
N/A	chr15	26044049	26044051	cg02334109	-1.6484	-6.2108	3.84E-06	0.002688417	0.7950	-0.2420	0.7527	0.7660	0.2411
NR4A2	chr2	157184879	157184881	cg00240195	-1.9769	-5.3095	2.99E-05	0.007299036	0.8686	-0.2419	0.8187	0.8679	0.3595
N/A	chr7	53393695	53393697	cg24859602	-1.6679	-6.4007	2.52E-06	0.002177049	0.8020	-0.2416	0.7700	0.7558	0.8165
IRF7	chr11	617139	617141	cg00645579	-1.5273	-6.2811	3.28E-06	0.002459659	0.7498	-0.2401	0.7075	0.7749	0.3611
N/A	chr12	56036795	56036797	cg23600372	-1.7521	-5.5139	1.86E-05	0.005804732	0.8296	-0.2385	0.8167	0.7947	0.3956
CAPN14 / EHD3	chr2	31455442	31455444	cg13149833	-1.4026	-6.0776	5.17E-06	0.003094249	0.6537	-0.2371	0.6828	0.6542	0.4075
COMMD3	chr10	22608673	22608675	cg01077178	-2.7482	-7.2019	4.50E-07	0.000916574	0.9414	-0.2362	0.8959	0.9301	0.8630
FBLN2	chr3	13679635	13679637	cg17054708	-1.6969	-5.0329	5.70E-05	0.009850227	0.8200	-0.2358	0.6742	0.8244	0.2287

Appendix 4: Top differentially methylated sites identified by genome-wide methylation profiling

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
WIPF1	chr2	175545837	175545839	cg00405190	-1.4481	-5.3363	2.81E-05	0.007105168	0.7190	-0.2351	0.8438	0.8103	0.4334
CUX1	chr7	101461822	101461824	cg12537379	-1.7067	-6.0426	5.59E-06	0.003193824	0.8237	-0.2350	0.8758	0.8485	0.6744
CBFA2T3	chr16	89024556	89024558	cg01534572	-1.3830	-6.2443	3.56E-06	0.00258875	0.6004	-0.2349	0.6029	0.6401	0.0536
TMC6	chr17	76117766	76117768	cg03297901	-2.0061	-5.1954	3.90E-05	0.00823586	0.8793	-0.2347	0.7279	0.8159	0.0789
N/A	chr3	154737056	154737058	cg06168950	-1.7040	-7.1226	5.32E-07	0.000995114	0.8238	-0.2345	0.8006	0.8677	0.5395
N/A	chr3	193629644	193629646	cg27203560	-1.4038	-6.1378	4.52E-06	0.002941399	0.6817	-0.2344	0.6396	0.6841	0.2402
N/A	chr6	1310484	1310486	cg17842822	-1.3866	-5.3205	2.91E-05	0.007215339	0.5787	-0.2343	0.5991	0.5746	0.2330
AL355531.2	chr10	131309401	131309403	cg00198994	-1.4388	-5.1523	4.31E-05	0.008601047	0.7160	-0.2341	0.7482	0.7339	0.7471
USP39 / C2orf68	chr2	85839649	85839651	cg04738827	-1.8310	-5.8737	8.18E-06	0.003851778	0.8520	-0.2340	0.8507	0.7822	0.8090
HOXA9	chr7	27209196	27209198	cg15912800	-2.3449	-5.2015	3.84E-05	0.008195212	0.3176	-0.2337	0.2929	0.3142	0.0548
N/A	chr4	183795784	183795786	cg12835012	-1.8386	-6.2687	3.38E-06	0.002491702	0.8540	-0.2335	0.7509	0.8254	0.2949
CD93	chr20	23067155	23067157	cg20438277	-1.3686	-5.4027	2.41E-05	0.006608799	0.6267	-0.2327	0.5078	0.4557	0.0553
WRB	chr21	40756695	40756697	cg07777652	-1.3735	-6.1881	4.04E-06	0.002767446	0.5810	-0.2324	0.5043	0.6280	0.2700
N/A	chr17	10075024	10075026	cg12245236	-1.7941	-5.9524	6.85E-06	0.003567484	0.3840	-0.2316	0.2973	0.1756	0.0880
HK2	chr2	75059601	75059603	cg18638581	-1.4100	-5.2567	3.38E-05	0.007687167	0.7071	-0.2311	0.6388	0.7487	0.3457
CLCN7	chr16	1521616	1521618	cg09050670	-2.4109	-5.0532	5.43E-05	0.009609233	0.9226	-0.2310	0.8976	0.9207	0.1894
N/A	chr5	2537519	2537521	cg00055986	-1.3616	-6.6810	1.37E-06	0.001604759	0.5691	-0.2296	0.3644	0.4321	0.3135
N/A	chr9	101011717	101011719	cg13709496	-1.5243	-5.3483	2.73E-05	0.006987768	0.7776	-0.2289	0.7882	0.7707	0.3856
DOCK10	chr2	225813193	225813195	cg20992319	-1.8133	-5.2090	3.77E-05	0.008134651	0.8545	-0.2289	0.8539	0.8686	0.8647
IFNGR2	chr21	34774163	34774165	cg08173915	-1.5958	-5.0873	5.02E-05	0.009229273	0.8036	-0.2285	0.8308	0.8409	0.3923
N/A	chr7	157475736	157475738	cg03983213	-1.4353	-5.3536	2.70E-05	0.006986219	0.7362	-0.2283	0.5509	0.7117	0.0967
MIR143 miRNA	chr5	148808455	148808457	cg16684117	-1.9963	-5.4985	1.93E-05	0.005878605	0.8842	-0.2274	0.8564	0.9277	0.3979
QRFP	chr9	133768965	133768967	cg03924115	-1.7808	-5.7023	1.21E-05	0.004615908	0.8503	-0.2272	0.8416	0.8531	0.4690
N/A	chr11	970663	970665	cg06064525	-2.0093	-5.2608	3.35E-05	0.007670993	0.8861	-0.2271	0.8517	0.9021	0.2674
N/A	chr11	134254645	134254647	cg09447621	-1.8041	-6.0957	4.96E-06	0.003051182	0.8551	-0.2268	0.7778	0.8349	0.2436
N/A	chr4	174431729	174431731	cg09867290	-1.5351	-5.1147	4.71E-05	0.008950337	0.7873	-0.2264	0.5910	0.4572	0.3391
N/A	chr8	1106400	1106402	cg03907855	-2.1418	-6.1636	4.26E-06	0.002850094	0.9020	-0.2261	0.9175	0.8858	0.9526

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

For all results tabulated below, selection criteria pertaining to control tissue (duodenal and proximal stomach) were NOT applied. Expression averages for these specimens are shown for interest only. Filtering of adjusted p-value < 0.05 and logfc > |log2(1.5)| was applied to tabulated results below.

Table 1: Top 100 up- and down-regulated transcript clusters for differentiating intervention-requiring disease from normal squamous mucosa

3,799 up-regulated and 2,998 down-regulated transcript clusters were identified as differentially expressed between normal squamous epithelium and intervention requiring disease (N v HGD-EAC, adjusted p-value < 0.05 and logfc > log2(1.5). Tx: transcript, logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, Exp_{base}: average expression in baseline samples, here, normal squamous epithelium, Exp_{duo}: average expression in duodenal epithelium, Exp_{stom}: average expression in proximal stomach epithelium, str: strand. Top 100 up- and down-regulated transcript clusters based on greatest |logfc|; listed in order of decreasing |logfc|.

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
Top 100 up-regulated transcript clusters from the N v HGD-EAC comparison (differentiating intervention requiring disease from normal squamous epithelium)													
TC07001164.hg.1	6.3114	18.4387	2.41E-14	2.03E-10	2.4452	8.8125	7.8018	chr7	16899029	16921613	-	NM_176813	AGR3
TC03001719.hg.1	6.2779	33.0766	1.95E-19	1.32E-14	4.2162	10.9536	8.2634	chr3	124624289	124672663	-	NM_033049	MUC13
TC07002495.hg.1	5.5034	14.9081	1.47E-12	4.97E-09	3.5054	9.8782	5.9416	chr7	100677704	100680561	+	---	---
TC07002496.hg.1	5.4451	15.3308	8.64E-13	3.43E-09	3.7525	10.1808	6.0229	chr7	100678736	100679444	+	---	---
TC01003051.hg.1	5.4311	10.5644	8.36E-10	3.70E-07	3.9260	8.4103	5.9298	chr1	120336641	120354283	-	NM_001159352	REG4
TC05001911.hg.1	5.2287	13.0717	1.75E-11	2.75E-08	4.3968	10.3294	10.7604	chr5	147204131	147211349	-	NM_003122	SPINK1
TC04000836.hg.1	5.1237	5.9764	6.55E-06	1.55E-04	3.0216	8.2071	10.5192	chr4	169013666	169108893	+	NM_007193	ANXA10
TC19001505.hg.1	5.1184	20.1084	4.37E-15	7.32E-11	5.8753	11.3040	10.3407	chr19	39292311	39303740	-	NM_006149	LGALS4
TC0X001747.hg.1	5.1092	11.7019	1.34E-10	1.12E-07	4.0778	9.8024	6.0420	chrX	65486697	65487236	+	---	---
TC04002303.hg.1	5.0681	10.2300	1.47E-09	5.08E-07	3.7880	9.2823	5.8750	chr4	165675283	165724947	+	---	---
TC04001926.hg.1	5.0109	4.7156	1.21E-04	1.39E-03	2.6228	9.4718	9.7421	chr4	15969851	15970275	+	---	---
TC08001431.hg.1	4.9964	13.7593	6.72E-12	1.33E-08	4.2965	10.3560	4.6983	chr8	95139394	95229531	-	NM_001144663	CDH17
TC6_apd_hap1000067.hg.1	4.9435	13.6657	7.64E-12	1.47E-08	3.6191	8.0192	6.7959	chr6_apd_hap1	1318454	1321112	-	NM_007028	TRIM31
TC19001042.hg.1	4.8748	13.1419	1.59E-11	2.55E-08	4.3460	11.1106	10.1189	chr19	3474405	3480540	-	NM_001136503	C19orf77
TC02004392.hg.1	4.8563	5.7701	1.04E-05	2.21E-04	3.6442	11.8566	6.8967	chr2	88422556	88427635	-	---	---
TC10001754.hg.1	4.8432	13.4147	1.08E-11	1.92E-08	3.5286	9.1097	5.0902	chr10	129676105	129691211	-	NM_152311	CLRN3
TC11002237.hg.1	4.8177	12.7749	2.68E-11	3.63E-08	3.6472	4.1360	3.7047	chr11	102733464	102745764	-	NM_002426	MMP12
TC12001723.hg.1	4.7623	29.0987	2.64E-18	8.90E-14	3.3606	8.0494	8.1431	chr12	71518865	71835678	-	NM_004616	TSPAN8

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC07003086.hg.1	4.7237	13.2845	1.30E-11	2.19E-08	5.3527	11.0530	8.7097	chr7	100550661	100551637	-	---	---
TC04000821.hg.1	4.6434	9.2612	8.12E-09	1.48E-06	2.5396	7.7616	4.4132	chr4	165675216	165724947	+	NR_038834	LOC100505989
TC0X001299.hg.1	4.6139	5.3967	2.46E-05	4.19E-04	4.2781	4.2814	4.3001	chrX	115592849	115594164	-	NM_001017978	CXorf61
TC03002462.hg.1	4.5782	10.9093	4.73E-10	2.50E-07	3.3366	8.8719	8.1115	chr3	110917882	110920289	+	---	---
TC01001720.hg.1	4.5535	6.7337	1.23E-06	4.52E-05	5.4355	10.8512	11.4785	chr1	206317459	206332104	+	NM_001910	CTSE
TC16000645.hg.1	4.5416	12.2887	5.49E-11	6.28E-08	5.2593	11.0469	8.6785	chr16	82068858	82132139	+	NM_002153	HSD17B2
TC03003109.hg.1	4.5205	5.6178	1.48E-05	2.88E-04	2.9071	4.1455	9.6955	chr3	137749545	137750672	-	---	---
TC13000228.hg.1	4.4536	6.8064	1.06E-06	4.07E-05	4.5295	10.8360	5.5739	chr13	53602830	53626196	+	NM_006418	OLFM4
TC07000640.hg.1	4.4508	11.2858	2.58E-10	1.72E-07	4.2975	9.4392	5.6876	chr7	100663353	100702140	+	NM_001040105	MUC17
TC03000545.hg.1	4.4442	13.0340	1.85E-11	2.79E-08	3.2250	8.7148	6.4195	chr3	108015376	108097132	+	NM_007072	HHLA2
TC11002234.hg.1	4.3521	10.4597	9.96E-10	4.09E-07	3.3581	5.8857	5.4730	chr11	102660641	102668966	-	NM_001145938	MMP1
TC05002798.hg.1	4.3380	8.4887	3.45E-08	3.84E-06	5.0224	8.8223	9.0193	chr5	553527	554002	-	---	---
TC07002497.hg.1	4.3011	10.5666	8.33E-10	3.70E-07	4.3495	9.3144	5.5696	chr7	100701581	100701790	+	---	---
TC02002835.hg.1	4.2644	7.9473	9.95E-08	7.66E-06	3.0113	9.6065	3.5581	chr2	228226872	228246711	-	NM_024795	TM4SF20
TC21000942.hg.1	4.2271	7.9856	9.22E-08	7.26E-06	3.7456	5.4679	9.8031	chr21	36089937	36090503	-	---	---
TC01000580.hg.1	4.2243	19.8898	5.42E-15	7.32E-11	5.2139	8.4753	9.5359	chr1	46640745	46651634	+	NM_005727	TSPAN1
TC07002629.hg.1	4.2204	17.8501	4.55E-14	2.89E-10	3.3188	8.6340	3.7731	chr7	141870970	141923474	+	---	---
TC11000788.hg.1	4.2014	9.9769	2.27E-09	6.67E-07	4.6610	8.3474	7.1957	chr11	74166393	74167870	+	---	---
TC11003279.hg.1	4.1560	10.5107	9.14E-10	3.89E-07	5.3354	8.7481	10.2169	chr11	85416009	85420543	-	---	---
TC07001163.hg.1	4.1324	14.9666	1.37E-12	4.86E-09	3.5359	7.4353	7.7732	chr7	16831435	16873057	-	NM_006408	AGR2
TC02002478.hg.1	3.8579	11.9361	9.35E-11	9.28E-08	4.1374	9.1093	6.1959	chr2	162848751	162931052	-	NM_001935	DPP4
TC12001163.hg.1	3.8540	6.6704	1.41E-06	5.00E-05	3.3781	9.4316	5.6822	chr12	7801996	7818502	-	NM_001644	APOBEC1
TC05002656.hg.1	3.8370	12.9948	1.96E-11	2.79E-08	6.6027	10.6394	11.7181	chr5	139554736	139623371	+	---	---
TC06001495.hg.1	3.8286	11.2445	2.75E-10	1.74E-07	3.9586	7.3347	6.2694	chr6	30070674	30080883	-	NM_007028	TRIM31
TC07000104.hg.1	3.8114	8.6952	2.33E-08	2.98E-06	2.6294	6.4699	3.1402	chr7	15707572	15721605	+	OTTHUMT00000326059	OTTHUMG00000152391
TC07002158.hg.1	3.8077	7.9978	9.00E-08	7.17E-06	2.3500	6.1103	2.8304	chr7	15707572	15721605	+	---	---
TC05002612.hg.1	3.7994	14.7924	1.71E-12	5.24E-09	4.9517	8.7382	9.3989	chr5	127522490	127522778	+	---	---
TC15002310.hg.1	3.7584	7.1813	4.76E-07	2.25E-05	4.6755	6.7997	5.5247	chr15	80253413	80263461	+	---	---
TC01003498.hg.1	3.7502	9.5341	4.96E-09	1.09E-06	3.7164	5.2532	4.9231	chr1	169481192	169555826	-	NM_000130	F5
TC03001692.hg.1	3.7480	8.2863	5.10E-08	5.05E-06	3.6063	8.1441	8.1161	chr3	120347015	120401418	-	NM_000187	HGD
TC07000726.hg.1	3.7347	8.0810	7.63E-08	6.51E-06	3.9643	9.0968	4.8152	chr7	117105838	117308719	+	NM_000492	CFTR
TC12001909.hg.1	3.7328	14.1809	3.81E-12	8.86E-09	4.0970	8.6060	8.9983	chr12	105196331	105352522	-	NM_032148	SLC41A2
TC07002911.hg.1	3.6990	8.4343	3.83E-08	4.10E-06	3.6934	5.5731	4.3933	chr7	41724713	41726643	-	---	---
TC03001966.hg.1	3.6905	5.5899	1.58E-05	3.03E-04	2.7781	10.9035	4.2462	chr3	164696686	164796283	-	NM_001041	SI
TC07001607.hg.1	3.6884	6.7566	1.17E-06	4.38E-05	2.9887	7.4057	5.9401	chr7	92817899	92855837	-	NM_001039372	HEPACAM2
TC06000779.hg.1	3.6845	11.9156	9.65E-11	9.31E-08	4.4808	8.5530	6.8147	chr6	86159302	86205509	+	NM_001204813	NT5E
TC02000663.hg.1	3.6792	6.0664	5.35E-06	1.33E-04	4.0127	7.3808	9.3863	chr2	108905095	108926371	+	NM_001056	SULT1C2

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC01000939.hg.1	3.6702	11.9581	9.04E-11	9.11E-08	4.7383	8.5278	9.1496	chr1	109656585	109749403	+	NM_020775	KIAA1324
TC11000835.hg.1	3.6645	9.1497	9.96E-09	1.68E-06	4.5680	6.9778	8.3597	chr11	77726761	77727541	+	---	---
TC6_qbl_hap6000121.hg.1	3.6590	10.9808	4.21E-10	2.40E-07	4.3113	7.5305	6.4799	chr6_qbl_hap6	1363640	1373843	-	NM_007028	TRIM31
TC04000408.hg.1	3.6558	6.4473	2.30E-06	7.12E-05	3.4989	3.5081	3.2133	chr4	74606223	74609433	+	NM_000584	IL8
TC6_mann_hap4000109.hg.1	3.6241	10.8830	4.94E-10	2.58E-07	4.3768	7.5584	6.4977	chr6_mann_hap4	1369234	1379436	-	NM_007028	TRIM31
TC06003087.hg.1	3.6236	14.2657	3.40E-12	8.83E-09	2.8769	6.0188	7.3205	chr6	138661322	138665748	+	---	---
TC11002235.hg.1	3.5996	6.8195	1.03E-06	3.99E-05	3.2416	4.3371	3.9633	chr11	102706528	102714534	-	NM_002422	MMP3
TC6_ssto_hap7000112.hg.1	3.5946	10.9546	4.39E-10	2.42E-07	4.3653	7.4995	6.4556	chr6_ssto_hap7	1401127	1411327	-	NM_007028	TRIM31
TC6_mcf_hap5000111.hg.1	3.5681	10.8564	5.16E-10	2.63E-07	4.3540	7.4624	6.4218	chr6_mcf_hap5	1452557	1462764	-	NM_007028	TRIM31
TC02004744.hg.1	3.5678	8.9452	1.45E-08	2.19E-06	4.4454	8.6138	6.0504	chr2	188328960	188330769	-	---	---
TC17001473.hg.1	3.5675	4.1583	4.55E-04	3.75E-03	4.0590	10.7581	8.9311	chr17	39030852	39041495	-	NM_019010	KRT20
TC07000902.hg.1	3.5670	16.0853	3.43E-13	1.82E-09	3.3879	7.9542	3.5929	chr7	141811549	141921088	+	NR_003715	LOC93432
TC02000488.hg.1	3.5611	8.5415	3.12E-08	3.57E-06	4.0763	8.8376	7.7078	chr2	79347488	79350545	+	NM_002909	REG1A
TC12001803.hg.1	3.5578	9.0718	1.15E-08	1.84E-06	4.4349	8.5463	7.7317	chr12	91496406	91505608	-	NM_002345	LUM
TC6_apd_hap1000090.hg.1	3.5271	11.7823	1.18E-10	1.04E-07	5.5961	9.0822	9.0774	chr6_apd_hap1	3145724	3161577	-	NM_001178044	SLC44A4
TC6_ssto_hap7000145.hg.1	3.5271	11.7823	1.18E-10	1.04E-07	5.5961	9.0822	9.0774	chr6_ssto_hap7	3163717	3179566	-	NM_001178044	SLC44A4
TC6_mcf_hap5000151.hg.1	3.5269	11.7713	1.20E-10	1.04E-07	5.5958	9.0829	9.0767	chr6_mcf_hap5	3210847	3226696	-	NM_001178044	SLC44A4
TC21000486.hg.1	3.4950	5.1505	4.36E-05	6.39E-04	6.7498	10.4136	12.1854	chr21	43782391	43786703	-	NM_003225	TFF1
TC6_dbb_hap3000120.hg.1	3.4793	10.4666	9.85E-10	4.08E-07	4.4217	7.4503	6.4023	chr6_dbb_hap3	1364399	1374601	-	NM_007028	TRIM31
TC11002228.hg.1	3.4767	6.6162	1.59E-06	5.46E-05	4.3612	5.4140	4.4952	chr11	102391239	102401484	-	NM_002423	MMP7
TC6_qbl_hap6000166.hg.1	3.4762	11.7919	1.17E-10	1.04E-07	5.6029	9.0342	9.0051	chr6_qbl_hap6	3124760	3140615	-	NM_001178044	SLC44A4
TC04002588.hg.1	3.4513	5.1044	4.85E-05	6.91E-04	2.7599	3.0278	2.8318	chr4	74861359	74862455	-	---	---
TC08000499.hg.1	3.4454	13.2963	1.28E-11	2.19E-08	3.8938	8.6670	6.3749	chr8	76320149	76479078	+	NM_004133	HNF4G
TC6_dbb_hap3000165.hg.1	3.4423	11.5834	1.61E-10	1.25E-07	5.6097	9.0117	8.9988	chr6_dbb_hap3	3116555	3132405	-	NM_001178044	SLC44A4
TC06001549.hg.1	3.4417	11.5920	1.59E-10	1.25E-07	5.6096	9.0110	8.9991	chr6	31830969	31846823	-	NM_001178044	SLC44A4
TC6_cox_hap2000176.hg.1	3.4417	11.5920	1.59E-10	1.25E-07	5.6096	9.0110	8.9991	chr6_cox_hap2	3340716	3356577	-	NM_001178044	SLC44A4
TC03000806.hg.1	3.4410	5.4696	2.08E-05	3.71E-04	4.4273	10.3388	8.0240	chr3	149191761	149221181	+	NM_004617	TM4SF4
TC05001556.hg.1	3.4168	6.0179	5.96E-06	1.44E-04	3.1773	6.1503	6.6863	chr5	82837296	82877139	-	ENST0000051389	VCAN-AS1

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
												9	
TC0X001610.hg.1	3.4159	10.5927	7.97E-10	3.64E-07	3.6586	6.4117	7.7453	chrX	17168618	17171104	+	---	---
TC06002119.hg.1	3.3936	6.5879	1.69E-06	5.70E-05	3.7363	9.9695	3.9040	chr6	133001997	133035194	-	NM_004666	VNN1
TC15001539.hg.1	3.3718	5.7748	1.03E-05	2.19E-04	4.6285	8.5096	8.2462	chr15	63362224	63364110	-	---	---
TC07001227.hg.1	3.3716	9.7872	3.17E-09	8.10E-07	3.1313	3.2578	2.9617	chr7	27233122	27239725	-	NM_000522	HOXA13
TC11002162.hg.1	3.3688	10.3833	1.13E-09	4.40E-07	4.7160	7.6027	9.1337	chr11	85405265	85522184	-	NM_001162951	SYTL2
TC12002508.hg.1	3.3559	3.7646	1.16E-03	7.49E-03	3.4514	10.2113	4.6968	chr12	92271167	92274184	+	---	---
TC17000050.hg.1	3.3429	7.1245	5.37E-07	2.48E-05	5.4769	10.9838	9.1509	chr17	4675187	4686508	+	NM_003963	TM4SF5
TC6_cox_hap2000131.hg.1	3.3404	9.8494	2.84E-09	7.52E-07	4.5392	7.4426	6.4162	chr6_cox_hap2	1582796	1593011	-	NM_007028	TRIM31
TC10000890.hg.1	3.3383	6.3617	2.78E-06	8.17E-05	5.2124	10.3478	5.2156	chr10	124320181	124403252	+	NM_004406	DMBT1
TC06003855.hg.1	3.3296	6.4106	2.49E-06	7.55E-05	3.4881	5.2683	3.8587	chr6	133065009	133079033	-	---	---
TC16000214.hg.1	3.3259	7.1763	4.82E-07	2.27E-05	5.0996	8.3160	7.3764	chr16	19421861	19510435	+	NM_001105248	TMC5
TC04001092.hg.1	3.3227	15.7995	4.85E-13	2.18E-09	4.9611	8.3840	8.2981	chr4	25749049	25865382	-	NM_015187	SEL1L3
TC12002823.hg.1	3.3160	11.3361	2.38E-10	1.62E-07	2.4148	6.6428	5.2070	chr12	31535160	31537109	-	---	---
TC03000878.hg.1	3.3148	5.3448	2.77E-05	4.59E-04	3.6472	8.3997	8.0774	chr3	160394947	160396240	+	NM_025047	ARL14
TC09000138.hg.1	3.3099	6.1175	4.77E-06	1.21E-04	6.2672	9.2864	6.1852	chr9	33218363	33248565	+	NM_014471	SPINK4
TC07002628.hg.1	3.3024	14.2896	3.29E-12	8.83E-09	3.4898	7.7724	3.3495	chr7	141811549	141843783	+	---	---
TC11003311.hg.1	3.2907	7.9369	1.02E-07	7.76E-06	4.8586	6.3095	5.9152	chr11	102667775	102668070	-	---	---
Top 100 down-regulated transcript clusters from the N v HGD-EAC comparison (differentiating intervention requiring disease from normal squamous epithelium)													
TC04001258.hg.1	-6.3568	-6.0146	6.01E-06	1.45E-04	10.7669	2.6350	2.6667	chr4	69092371	69111438	-	NM_182502	TMPRSS11B
TC04000369.hg.1	-6.3452	-6.5538	1.82E-06	6.01E-05	11.0680	3.4914	3.4121	chr4	69313167	69363322	+	NM_014058	TMPRSS11E
TC06001785.hg.1	-6.3423	-6.5305	1.92E-06	6.23E-05	11.0703	3.4573	3.2183	chr6	49695092	49712168	-	NM_001190986	CRISP3
TC4_ctg9_hap1000001.hg.1	-6.3325	-6.5117	2.00E-06	6.43E-05	11.0813	3.4932	3.3825	chr4_ctg9_hap1	143089	193241	+	NM_014058	TMPRSS11E
TC04002642.hg.1	-6.2743	-11.9266	9.49E-11	9.28E-08	11.4194	4.0165	8.4066	chr4	100340208	100341836	-	---	---
TC01000821.hg.1	-6.2403	-6.7377	1.22E-06	4.49E-05	10.2816	3.3565	2.9857	chr1	86889769	86922241	+	NM_006536	CLCA2
TC04001253.hg.1	-6.1784	-7.0108	6.83E-07	2.96E-05	9.5993	2.8341	2.7737	chr4	68775103	68829858	-	NM_001114387	TMPRSS11A
TC01000835.hg.1	-6.1105	-7.0146	6.77E-07	2.95E-05	10.5098	3.6973	3.8221	chr1	89829436	89853719	+	NM_198460	GBP6
TC02000716.hg.1	-6.0653	-7.0829	5.86E-07	2.65E-05	11.5806	4.5013	4.5807	chr2	113763038	113766126	+	BC107043	IL36A
TC18000228.hg.1	-5.9244	-6.8624	9.36E-07	3.72E-05	10.1920	3.3984	3.4266	chr18	61254223	61271873	+	NM_012397	SERPINB13
TC12000136.hg.1	-5.8960	-6.5996	1.65E-06	5.61E-05	10.9389	3.8276	3.8357	chr12	8975068	9039597	+	NM_144670	A2ML1
TC16000456.hg.1	-5.8725	-7.5739	2.11E-07	1.26E-05	10.0465	3.7904	3.6352	chr16	55600584	55601599	+	NM_032330	CAPNS2
TC01004511.hg.1	-5.8452	-5.8617	8.48E-06	1.89E-04	11.5816	3.6352	2.8749	chr1	87012759	87046432	+	---	---
TC18000124.hg.1	-5.8178	-5.6277	1.44E-05	2.84E-04	10.2571	2.9679	2.7953	chr18	28898052	28937393	+	NM_001942	DSG1
TC18000188.hg.1	-5.8122	-7.1980	4.60E-07	2.20E-05	8.8368	2.8528	2.7814	chr18	52258390	52266724	+	NM_173629	DYNAP
TC04001412.hg.1	-5.7854	-11.8940	9.97E-11	9.48E-08	9.6438	3.4076	6.6783	chr4	100333418	100356894	-	NM_000673	ADH7
TC04000413.hg.1	-5.6799	-7.0198	6.70E-07	2.93E-05	8.2143	2.3131	2.2276	chr4	75174190	75181024	+	NM_001013442	EPGN

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC01000823.hg.1	-5.6718	-5.9555	6.86E-06	1.61E-04	11.0913	3.5421	2.8544	chr1	87012759	87046437	+	NM_012128	CLCA4
TC11001505.hg.1	-5.6479	-8.2075	5.95E-08	5.60E-06	8.8387	3.0188	2.9702	chr11	26580579	26593815	-	NM_001135091	MUC15
TC01005829.hg.1	-5.6146	-5.5926	1.57E-05	3.02E-04	9.5017	3.4639	3.4949	chr1	152286184	152286725	-	---	---
TC6_mcf_hap5000038.hg.1	-5.5672	-6.5773	1.73E-06	5.80E-05	11.4233	4.8206	4.8134	chr6_mcf_hap5	2333431	2339631	+	NM_001010909	MUC21
TC6_apd_hap1000031.hg.1	-5.5497	-6.5418	1.87E-06	6.12E-05	11.4397	4.8544	4.8559	chr6_apd_hap1	2263124	2269314	+	NM_001010909	MUC21
TC6_dbb_hap3000044.hg.1	-5.5497	-6.5418	1.87E-06	6.12E-05	11.4397	4.8544	4.8559	chr6_dbb_hap3	2245596	2251786	+	NM_001010909	MUC21
TC04001255.hg.1	-5.4948	-7.6800	1.70E-07	1.10E-05	8.0668	2.4031	2.4842	chr4	68918916	68995598	-	NM_207407	TMPRSS11F
TC18000435.hg.1	-5.4481	-6.1788	4.16E-06	1.10E-04	10.1880	3.4078	3.0548	chr18	28570052	28622781	-	NM_001941	DSC3
TC01003247.hg.1	-5.4446	-6.2256	3.75E-06	1.02E-04	11.7091	5.0342	5.0450	chr1	152381719	152386750	-	NM_016190	CRNN
TC20000023.hg.1	-5.4441	-6.0622	5.40E-06	1.34E-04	11.6566	4.9902	4.8450	chr20	2276613	2321725	+	NM_003245	TGM3
TC04002578.hg.1	-5.4294	-7.4071	2.98E-07	1.61E-05	7.2778	1.7763	1.7791	chr4	69049896	69051921	-	---	---
TC04001257.hg.1	-5.3799	-7.7687	1.42E-07	9.58E-06	7.4605	1.8600	2.0223	chr4	69054242	69083798	-	NM_001129907	TMPRSS11BNL
TC13000295.hg.1	-5.2383	-5.8960	7.85E-06	1.78E-04	10.3551	2.7336	3.3748	chr13	78049967	78219398	+	NM_001160706	SCEL
TC19001572.hg.1	-5.2273	-6.9726	7.40E-07	3.13E-05	12.0359	6.1708	5.9529	chr19	42891171	42894444	-	NM_032488	CNFN
TC01001244.hg.1	-5.2004	-7.6408	1.84E-07	1.15E-05	9.4470	3.8030	3.9397	chr1	152881021	152884362	+	NM_005547	IVL
TC6_cox_hap2000051.hg.1	-5.1827	-6.2706	3.39E-06	9.49E-05	11.7406	5.2996	5.5784	chr6_cox_hap2	2463385	2469590	+	NM_001010909	MUC21
TC18001005.hg.1	-5.1361	-5.4747	2.05E-05	3.68E-04	10.3935	3.4489	3.5458	chr18	61322431	61329197	-	NM_006919	SERPINB3
TC01001488.hg.1	-5.1269	-6.8412	9.80E-07	3.86E-05	9.8172	4.2112	4.4631	chr1	171154347	171181822	+	NM_001460	FMO2
TC18001004.hg.1	-5.1073	-5.4743	2.06E-05	3.68E-04	9.7904	3.3744	3.3226	chr18	61304493	61311532	-	NM_002974	SERPINB4
TC04001252.hg.1	-5.0828	-5.6459	1.39E-05	2.75E-04	10.3312	3.9104	3.7742	chr4	68686594	68749750	-	NM_004262	TMPRSS11D
TC01005830.hg.1	-5.0365	-5.2383	3.55E-05	5.50E-04	7.3810	2.0343	2.2763	chr1	152287087	152287969	-	---	---
TC17001468.hg.1	-5.0223	-5.1302	4.57E-05	6.60E-04	9.2495	3.7841	3.5375	chr17	38854243	38860002	-	NM_019016	KRT24
TC6_mann_hap4000046.hg.1	-5.0000	-6.2110	3.87E-06	1.04E-04	11.7601	5.4994	5.7704	chr6_mann_hap4	2299571	2305861	+	NM_001010909	MUC21
TC05000810.hg.1	-4.9901	-7.2258	4.34E-07	2.10E-05	9.3987	3.6424	4.7522	chr5	147691982	147695485	+	NM_032566	SPINK7
TC15001829.hg.1	-4.9868	-7.3270	3.52E-07	1.81E-05	10.6656	5.1489	5.1443	chr15	89998680	90039844	-	NM_016321	RHCG
TC01006366.hg.1	-4.9742	-8.7526	2.09E-08	2.77E-06	9.0579	4.3205	4.4973	chr1	115125469	115213043	-	NM_001256404	DENND2C
TC19001446.hg.1	-4.9579	-6.1356	4.58E-06	1.18E-04	11.0953	4.9862	4.8912	chr19	36014269	36019253	-	NM_198538	SBSN
TC13001343.hg.1	-4.9389	-7.4007	3.02E-07	1.62E-05	9.5388	4.1725	4.2259	chr13	20797502	20806372	-	---	---
TC6_qbl_hap6000044.hg.1	-4.8630	-6.0500	5.55E-06	1.37E-04	11.7903	5.4797	5.9121	chr6_qbl_hap6	2244470	2250865	+	NM_001010909	MUC21
TC12001538.hg.1	-4.8538	-8.2918	5.05E-08	5.01E-06	10.7002	5.5556	5.4813	chr12	53231588	53242778	-	NM_173352	KRT78
TC02001715.hg.1	-4.8068	-9.9149	2.53E-09	7.09E-07	8.9278	3.9288	3.8544	chr2	31395922	31456724	-	NM_001145122	CAPN14
TC12001536.hg.1	-4.7999	-5.2816	3.21E-05	5.14E-04	11.9955	5.1747	5.0701	chr12	53200327	53208335	-	NM_002272	KRT4
TC06000353.hg.1	-4.7494	-6.3924	2.59E-06	7.75E-05	11.8250	6.0283	6.1762	chr6	30951485	30957680	+	NM_001010909	MUC21

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC12002789.hg.1	-4.7243	-6.2908	3.25E-06	9.19E-05	7.5141	2.8223	2.7429	chr12	21689123	21757554	-	---	---
TC06003208.hg.1	-4.6975	-7.4081	2.97E-07	1.61E-05	7.6543	2.7852	2.7239	chr6	168084231	168088891	+	---	---
TC03001905.hg.1	-4.6411	-8.3434	4.57E-08	4.64E-06	8.2332	3.5838	3.3951	chr3	151011876	151034740	-	NM_023915	GPR87
TC04002094.hg.1	-4.6041	-8.0889	7.52E-08	6.43E-06	9.6699	5.2431	6.6181	chr4	87515885	87736302	+	---	---
TC10002963.hg.1	-4.5910	-8.3065	4.91E-08	4.90E-06	10.0795	5.1799	5.0563	chr10	47157983	47174122	-	NM_001098845	ANXA8L1
TC04000282.hg.1	-4.5826	-8.0363	8.34E-08	6.90E-06	8.5796	4.4648	6.5740	chr4	48988264	49064098	+	NM_025087	CWH43
TC14000971.hg.1	-4.5679	-7.4266	2.86E-07	1.56E-05	9.9252	5.0235	4.9707	chr14	24718320	24732416	-	NM_000359	TGM1
TC19000538.hg.1	-4.5642	-8.1952	6.10E-08	5.68E-06	10.0305	5.6299	5.2216	chr19	39687601	39692524	+	NM_001001414	NCCRP1
TC18000127.hg.1	-4.5535	-6.7730	1.13E-06	4.27E-05	11.2379	6.7588	3.1241	chr18	29027732	29058665	+	NM_001944	DSG3
TC01005834.hg.1	-4.5038	-4.0986	5.24E-04	4.16E-03	11.2587	3.5787	3.0263	chr1	153012201	153013594	-	---	---
TC11002740.hg.1	-4.5006	-7.8738	1.15E-07	8.45E-06	10.8745	6.3786	5.9918	chr11	67429963	67430223	+	---	---
TC17000513.hg.1	-4.5006	-5.8977	7.82E-06	1.78E-04	11.1606	5.1270	4.9398	chr17	39657235	39658661	+	OTTHUMT00000257901	OTTHUMG00000133684
TC12001297.hg.1	-4.4970	-6.0998	4.96E-06	1.25E-04	7.0979	2.6419	2.5754	chr12	21689123	21757781	-	NM_021957	GYS2
TC01006378.hg.1	-4.4970	-4.8689	8.43E-05	1.06E-03	9.3667	3.3687	3.4908	chr1	153112594	153113969	-	NR_003062	SPRR2C
TC04000469.hg.1	-4.4889	-8.0617	7.93E-08	6.65E-06	9.2999	4.9455	6.2504	chr4	87515468	87736328	+	NM_006264	PTPN13
TC14000080.hg.1	-4.4693	-6.2460	3.58E-06	9.88E-05	9.1969	4.5054	4.4745	chr14	21510385	21512393	+	NM_032572	RNASE7
TC01005836.hg.1	-4.4523	-4.8884	8.05E-05	1.02E-03	9.2679	3.4210	3.6068	chr1	153112594	153113969	-	---	---
TC0X001579.hg.1	-4.4012	-7.5003	2.46E-07	1.40E-05	8.9929	3.9831	3.7414	chrX	2670337	2693037	+	---	---
TC10000314.hg.1	-4.3983	-8.0296	8.45E-08	6.93E-06	9.6810	4.9928	4.7928	chr10	48255225	48279199	+	NM_001098845	ANXA8L1
TC03002609.hg.1	-4.3949	-9.0807	1.13E-08	1.83E-06	8.7266	6.4090	5.0587	chr3	152557179	152559228	+	---	---
TC05000805.hg.1	-4.3779	-5.3968	2.46E-05	4.19E-04	11.1755	4.3385	5.7582	chr5	147405246	147516925	+	NM_001127699	SPINK5
TC01001254.hg.1	-4.3469	-6.5664	1.77E-06	5.89E-05	10.9544	5.9423	5.8387	chr1	153330330	153333503	+	NM_002965	S100A9
TC12001526.hg.1	-4.3334	-5.8414	8.88E-06	1.96E-04	11.5277	5.4670	5.3992	chr12	52908359	52914471	-	NM_000424	KRT5
TC01005828.hg.1	-4.3324	-5.2619	3.36E-05	5.28E-04	9.6127	4.8973	4.6607	chr1	152274640	152281460	-	---	---
TC10000309.hg.1	-4.3293	-8.0914	7.48E-08	6.43E-06	9.6987	5.1796	4.9404	chr10	47742393	47770871	+	NM_001630	ANXA8L2
TC01004997.hg.1	-4.3018	-7.0731	5.98E-07	2.69E-05	8.7065	4.2472	4.1867	chr1	209602165	209606183	+	---	---
TC06001902.hg.1	-4.2689	-7.4680	2.63E-07	1.47E-05	8.4768	4.4059	4.6807	chr6	80624529	80657315	-	NM_022726	ELOVL4
TC01001199.hg.1	-4.2679	-8.5456	3.09E-08	3.55E-06	8.4876	4.1976	4.1870	chr1	151009029	151020076	+	NM_001159642	BNIPL
TC03001042.hg.1	-4.2646	-7.5996	2.01E-07	1.22E-05	8.7630	4.3226	4.2599	chr3	189349205	189615068	+	NM_003722	TP63
TC06004001.hg.1	-4.2645	-7.3546	3.32E-07	1.74E-05	8.4543	3.8965	3.6994	chr6	168080306	168096970	-	---	---
TC12002888.hg.1	-4.2634	-5.2968	3.10E-05	5.01E-04	12.1279	6.0879	6.1057	chr12	52912731	52912944	-	---	---
TC04001254.hg.1	-4.2328	-5.9372	7.15E-06	1.66E-04	6.8616	2.6571	2.4297	chr4	68857530	68863205	-	NR_033737	TMPRSS11GP
TC03000836.hg.1	-4.2097	-9.3658	6.72E-09	1.31E-06	8.8243	6.4595	5.1308	chr3	152557179	152559228	+	OTTHUMT00000356944	OTTHUMG00000159695
TC10001522.hg.1	-4.1915	-8.8850	1.63E-08	2.35E-06	9.3831	5.4160	6.8116	chr10	93388197	93392858	-	NM_005398	PPP1R3C
TC12001436.hg.1	-4.1846	-8.3825	4.23E-08	4.40E-06	8.0995	3.9126	3.8091	chr12	48103517	48119355	-	NM_001172439	ENDOU
TC06002317.hg.1	-4.1651	-7.3481	3.37E-07	1.76E-05	7.8980	3.4924	3.2863	chr6	168080306	168096970	-	BC093745	LOC441178

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC08001637.hg.1	-4.1503	-6.8333	9.96E-07	3.91E-05	8.1732	3.4877	3.4292	chr8	130760442	130799134	-	NM_031415	GSDMC
TC6_mann_hap4000176.hg.1	-4.1104	-6.3103	3.11E-06	8.95E-05	6.8706	3.0106	2.5619	chr6_mann_hap4	2370339	2376003	+	---	---
TC01003245.hg.1	-4.1052	-5.2983	3.09E-05	5.00E-04	8.5067	4.1574	4.1342	chr1	152274651	152297679	-	NM_002016	FLG
TC21000361.hg.1	-4.0956	-5.3705	2.61E-05	4.39E-04	6.9501	2.6219	2.7697	chr21	31537882	31539074	-	ENST00000286808	CLDN17
TC01001747.hg.1	-4.0924	-6.9911	7.12E-07	3.04E-05	8.5964	4.3827	4.3584	chr1	209602165	209606183	+	NR_029622	MIR205
TC19001152.hg.1	-4.0530	-8.7028	2.29E-08	2.95E-06	8.1629	3.9707	4.0935	chr19	9800600	9811493	-	NM_001199814	ZNF812
TC18000227.hg.1	-4.0394	-6.2900	3.25E-06	9.20E-05	6.8332	2.8412	2.7312	chr18	61223393	61234244	+	NM_080474	SERPINB12
TC6_mcf_hap5000187.hg.1	-4.0372	-6.0127	6.03E-06	1.45E-04	6.9908	3.0713	2.7608	chr6_mcf_hap5	2403917	2409577	+	---	---
TC04002559.hg.1	-4.0226	-8.4004	4.09E-08	4.28E-06	10.5837	6.8799	7.1823	chr4	57514710	57522067	-	---	---
TC15001996.hg.1	-4.0045	-8.5064	3.33E-08	3.76E-06	7.8457	3.7938	3.6918	chr15	100940600	101084925	-	NM_178842	CERS3
TC12001524.hg.1	-4.0034	-5.0441	5.59E-05	7.72E-04	11.7605	6.2612	6.0276	chr12	52862300	52867569	-	NM_173086	KRT6C
TC21000282.hg.1	-3.9703	-6.9460	7.83E-07	3.27E-05	7.5708	4.4694	3.9218	chr21	15340810	15340916	-	---	---
TC06003648.hg.1	-3.9599	-7.5036	2.44E-07	1.40E-05	8.8855	4.4670	5.8041	chr6	42059976	42061997	-	---	---
TC06003638.hg.1	-3.9568	-8.0087	8.81E-08	7.05E-06	6.5756	2.3900	2.4994	chr6	40825508	40827706	-	---	---

Table 2: Top 100 up- and down-regulated transcript clusters for differentiating non-dysplastic Barrett's esophagus from normal mucosa

2,512 up-regulated and 1,984 down-regulated transcript clusters were identified as differentially expressed between normal squamous epithelium and non-dysplastic Barrett's epithelium (N v BE, adjusted p-value < 0.05 and logfc > log2(1.5). Tx: transcript, logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, Exp_{base}: average expression in baseline samples, here, normal squamous epithelium, Exp_{duo}: average expression in duodenal epithelium, Exp_{stom}: average expression in proximal stomach epithelium, str: strand. Top 100 up- and down-regulated transcript clusters based on greatest |logfc|; listed in order of decreasing |logfc|.

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
Top 100 up-regulated transcript clusters from the N v BE comparison (differentiating Barrett's mucosa from normal squamous epithelium)													
TC02004392.hg.1	7.6692	9.1123	1.07E-08	4.84E-06	3.6442	11.8566	6.8967	chr2	88422556	88427635	-	---	---
TC04001926.hg.1	7.3129	6.8820	8.98E-07	1.07E-04	2.6228	9.4718	9.7421	chr4	15969851	15970275	+	---	---
TC04000836.hg.1	6.9995	8.1643	6.48E-08	1.81E-05	3.0216	8.2071	10.5192	chr4	169013666	169108893	+	NM_007193	ANXA10
TC03001966.hg.1	6.8452	10.3680	1.16E-09	9.02E-07	2.7781	10.9035	4.2462	chr3	164696686	164796283	-	NM_001041	SI
TC17001473.hg.1	6.8321	7.9637	9.63E-08	2.37E-05	4.0590	10.7581	8.9311	chr17	39030852	39041495	-	NM_019010	KRT20
TC01003051.hg.1	6.8151	13.2566	1.35E-11	2.47E-08	3.9260	8.4103	5.9298	chr1	120336641	120354283	-	NM_001159352	REG4
TC03001719.hg.1	6.6770	35.1794	5.54E-20	3.74E-15	4.2162	10.9536	8.2634	chr3	124624289	124672663	-	NM_033049	MUC13
TC07001746.hg.1	6.5292	15.7106	5.40E-13	2.15E-09	3.2495	11.3912	5.1625	chr7	107405912	107443678	-	NM_000111	SLC26A3

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC07001164.hg.1	6.3814	18.6433	1.94E-14	1.87E-10	2.4452	8.8125	7.8018	chr7	16899029	16921613	-	NM_176813	AGR3
TC07002496.hg.1	6.1978	17.4501	7.08E-14	5.31E-10	3.7525	10.1808	6.0229	chr7	100678736	100679444	+	---	---
TC07002495.hg.1	6.1425	16.6395	1.78E-13	1.00E-09	3.5054	9.8782	5.9416	chr7	100677704	100680561	+	---	---
TC13000228.hg.1	6.1392	9.3826	6.52E-09	3.52E-06	4.5295	10.8360	5.5739	chr13	53602830	53626196	+	NM_006418	OLFM4
TC02002835.hg.1	5.8781	10.9545	4.39E-10	4.43E-07	3.0113	9.6065	3.5581	chr2	228226872	228246711	-	NM_024795	TM4SF20
TC08001431.hg.1	5.8277	16.0487	3.59E-13	1.61E-09	4.2965	10.3560	4.6983	chr8	95139394	95229531	-	NM_001144663	CDH17
TC01001720.hg.1	5.7680	8.5297	3.19E-08	1.09E-05	5.4355	10.8512	11.4785	chr1	206317459	206332104	+	NM_001910	CTSE
TC02000663.hg.1	5.5000	9.0686	1.16E-08	5.04E-06	4.0127	7.3808	9.3863	chr2	108905095	108926371	+	NM_001056	SULT1C2
TC12002508.hg.1	5.4797	6.1470	4.47E-06	3.01E-04	3.4514	10.2113	4.6968	chr12	92271167	92274184	+	---	---
TC12000709.hg.1	5.4769	6.0731	5.27E-06	3.33E-04	3.4963	10.2307	4.4731	chr12	92271167	92274184	+	---	---
TC0Y000083.hg.1	5.4480	5.6504	1.37E-05	6.11E-04	2.7882	2.7436	6.0001	chrY	25365581	25437503	+	NM_001005785	DAZ2
TC04001500.hg.1	5.4478	9.0957	1.10E-08	4.95E-06	2.3328	9.5233	2.8036	chr4	120238405	120243545	-	NM_000134	FABP2
TC03000806.hg.1	5.4187	8.6131	2.72E-08	9.46E-06	4.4273	10.3388	8.0240	chr3	149191761	149221181	+	NM_004617	TM4SF4
TC03003109.hg.1	5.4013	6.7124	1.29E-06	1.33E-04	2.9071	4.1455	9.6955	chr3	137749545	137750672	-	---	---
TC19001505.hg.1	5.3819	21.1434	1.61E-15	2.18E-11	5.8753	11.3040	10.3407	chr19	39292311	39303740	-	NM_006149	LGALS4
TC0X000177.hg.1	5.3282	11.6424	1.47E-10	1.65E-07	2.9150	9.6800	4.6208	chrX	38211736	38280703	+	NM_000531	OTC
TC21000486.hg.1	5.3163	7.8344	1.25E-07	2.80E-05	6.7498	10.4136	12.1854	chr21	43782391	43786703	-	NM_003225	TFF1
TC0X001747.hg.1	5.3010	12.1413	6.85E-11	8.26E-08	4.0778	9.8024	6.0420	chrX	65486697	65487236	+	---	---
TC06000631.hg.1	5.2950	10.1277	1.75E-09	1.24E-06	3.3383	9.3989	3.7220	chr6	46761094	46807519	+	NM_005588	MEP1A
TC0X000529.hg.1	5.2920	7.6756	1.72E-07	3.50E-05	4.2925	6.1041	10.3953	chrX	107288200	107322414	+	NM_001170553	VSIG1
TC15001837.hg.1	5.2196	16.3525	2.50E-13	1.20E-09	5.3104	11.3706	5.8109	chr15	90328120	90358094	-	NM_001150	ANPEP
TC04002303.hg.1	5.2165	10.5296	8.86E-10	7.29E-07	3.7880	9.2823	5.8750	chr4	165675283	165724947	+	---	---
TC01000822.hg.1	5.1943	9.2976	7.60E-09	3.83E-06	3.1142	10.1797	3.4495	chr1	86934051	86965974	+	NM_001285	CLCA1
TC05001911.hg.1	5.1628	12.9069	2.22E-11	3.57E-08	4.3968	10.3294	10.7604	chr5	147204131	147211349	-	NM_003122	SPINK1
TC07000640.hg.1	5.1472	13.0518	1.80E-11	3.12E-08	4.2975	9.4392	5.6876	chr7	100663353	100702140	+	NM_001040105	MUC17
TC07003086.hg.1	5.1434	14.4651	2.61E-12	7.68E-09	5.3527	11.0530	8.7097	chr7	100550661	100551637	-	---	---
TC0Y000093.hg.1	5.1163	5.6385	1.41E-05	6.21E-04	2.9781	2.8803	5.2257	chrY	26979967	27053187	+	NM_001005375	DAZ4
TC03001985.hg.1	5.1142	12.3569	4.96E-11	6.20E-08	3.0163	6.8658	6.5837	chr3	169556967	169587718	-	NM_024727	LRRC31
TC19001042.hg.1	5.0942	13.7335	6.96E-12	1.67E-08	4.3460	11.1106	10.1189	chr19	3474405	3480540	-	NM_001136503	C19orf77
TC16000645.hg.1	5.0671	13.7107	7.18E-12	1.67E-08	5.2593	11.0469	8.6785	chr16	82068858	82132139	+	NM_002153	HSD17B2
TC6_apd_hap1000067.hg.1	5.0321	13.9107	5.47E-12	1.37E-08	3.6191	8.0192	6.7959	chr6_apd_hap1	1318454	1321112	-	NM_007028	TRIM31
TC09001421.hg.1	5.0198	8.8234	1.83E-08	7.01E-06	4.7877	11.7273	8.7092	chr9	104182842	104198105	-	NM_000035	ALDOB
TC10000627.hg.1	4.9838	5.5424	1.76E-05	7.21E-04	3.6520	8.1987	10.4879	chr10	90424146	90438572	+	NM_001198828	LIPF
TC01003768.hg.1	4.9506	10.2001	1.55E-09	1.15E-06	5.2061	10.7348	7.9139	chr1	207101863	207119811	-	NM_002644	PIGR
TC01000580.hg.1	4.9159	23.1459	2.66E-16	4.48E-12	5.2139	8.4753	9.5359	chr1	46640745	46651634	+	NM_005727	TSPAN1
TC11003279.hg.1	4.8994	12.3909	4.71E-11	6.01E-08	5.3354	8.7481	10.2169	chr11	85416009	85420543	-	---	---
TC0Y000210.hg.1	4.8712	5.5490	1.73E-05	7.14E-04	3.0026	2.9094	5.2087	chrY	26909216	26959639	-	NM_020364	DAZ3

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC01003409.hg.1	4.8424	7.6855	1.68E-07	3.46E-05	4.3600	9.7169	6.2701	chr1	160846329	160854960	-	NM_017625	ITLN1
TC07002497.hg.1	4.8171	11.8341	1.09E-10	1.27E-07	4.3495	9.3144	5.5696	chr7	100701581	100701790	+	---	---
TC18000135.hg.1	4.8005	10.2362	1.45E-09	1.09E-06	3.1120	10.7710	3.6477	chr18	29769987	29800366	+	NM_005925	MEP1B
TC04000821.hg.1	4.7844	9.5425	4.89E-09	2.77E-06	2.5396	7.7616	4.4132	chr4	165675216	165724947	+	NR_038834	LOC100505989
TC03002462.hg.1	4.7791	11.3878	2.19E-10	2.28E-07	3.3366	8.8719	8.1115	chr3	110917882	110920289	+	---	---
TC10001754.hg.1	4.7720	13.2174	1.43E-11	2.53E-08	3.5286	9.1097	5.0902	chr10	129676105	129691211	-	NM_152311	CLRN3
TC06002119.hg.1	4.7505	9.2220	8.72E-09	4.24E-06	3.7363	9.9695	3.9040	chr6	133001997	133035194	-	NM_004666	VNN1
TC12001723.hg.1	4.7238	28.8638	3.11E-18	1.05E-13	3.3606	8.0494	8.1431	chr12	71518865	71835678	-	NM_004616	TSPAN8
TC0Y000196.hg.1	4.5882	5.5904	1.57E-05	6.67E-04	3.1479	3.0561	4.4589	chrY	25275502	25345254	-	NM_001005375	DAZ4
TC03000545.hg.1	4.5199	13.2559	1.35E-11	2.47E-08	3.2250	8.7148	6.4195	chr3	108015376	108097132	+	NM_007072	HHLA2
TC02003394.hg.1	4.4886	5.6237	1.46E-05	6.35E-04	4.7806	10.7495	9.5061	chr2	89185097	89185706	+	---	---
TC12001163.hg.1	4.4101	7.6328	1.88E-07	3.74E-05	3.3781	9.4316	5.6822	chr12	7801996	7818502	-	NM_001644	APOBEC1
TC03000510.hg.1	4.4048	8.3294	4.69E-08	1.44E-05	2.8646	8.7313	3.0343	chr3	100328433	100414323	+	NM_032787	GPR128
TC02002478.hg.1	4.3996	13.6121	8.22E-12	1.80E-08	4.1374	9.1093	6.1959	chr2	162848751	162931052	-	NM_001935	DPP4
TC02002069.hg.1	4.3689	8.2764	5.20E-08	1.54E-05	4.2701	9.2833	5.0923	chr2	88422508	88427650	-	NM_001443	FABP1
TC07000104.hg.1	4.3684	9.9659	2.32E-09	1.58E-06	2.6294	6.4699	3.1402	chr7	15707572	15721605	+	OTTHUMT00000326059	OTTHUMG00000152391
TC03001692.hg.1	4.3654	9.6511	4.03E-09	2.39E-06	3.6063	8.1441	8.1161	chr3	120347015	120401418	-	NM_000187	HGD
TC04001406.hg.1	4.3406	8.0685	7.82E-08	2.06E-05	3.6357	9.5860	3.7062	chr4	100044808	100078949	-	NM_000670	ADH4
TC07002629.hg.1	4.3321	18.3222	2.73E-14	2.30E-10	3.3188	8.6340	3.7731	chr7	141870970	141923474	+	---	---
TC07000726.hg.1	4.3270	9.3626	6.76E-09	3.59E-06	3.9643	9.0968	4.8152	chr7	117105838	117308719	+	NM_000492	CFTR
TC07001163.hg.1	4.3246	15.6626	5.73E-13	2.15E-09	3.5359	7.4353	7.7732	chr7	16831435	16873057	-	NM_006408	AGR2
TC07002158.hg.1	4.3189	9.0715	1.15E-08	5.04E-06	2.3500	6.1103	2.8304	chr7	15707572	15721605	+	---	---
TC08000907.hg.1	4.3169	5.2201	3.70E-05	1.18E-03	2.6901	11.1934	3.6942	chr8	6782215	6783598	-	NM_001926	DEFA6
TC02004093.hg.1	4.3083	7.4999	2.46E-07	4.44E-05	2.3753	10.9422	4.8379	chr2	21225293	21225747	-	---	---
TC01001831.hg.1	4.2953	7.7115	1.60E-07	3.32E-05	3.5952	7.0827	8.4796	chr1	222910549	222924147	+	NM_207468	FAM177B
TC04001261.hg.1	4.2692	6.1914	4.05E-06	2.83E-04	3.2448	8.6020	8.1218	chr4	69512315	69536494	-	NM_001076	UGT2B15
TC0Y000279.hg.1	4.2582	5.6256	1.45E-05	6.34E-04	3.4712	3.3489	4.2141	chrY	26980079	26999579	+	---	---
TC10002643.hg.1	4.2408	7.4980	2.47E-07	4.45E-05	2.5628	8.0562	4.5686	chr10	52559793	52565728	-	---	---
TC4_ctg9_hap1000004.hg.1	4.2237	6.1691	4.25E-06	2.92E-04	3.6915	9.5171	8.5147	chr4_ctg9_hap1	224704	249081	-	NM_001076	UGT2B15
TC11002492.hg.1	4.1910	9.0824	1.13E-08	5.01E-06	7.0391	11.3513	11.6740	chr11	1017314	1018174	+	---	---
TC11002162.hg.1	4.1660	12.8406	2.44E-11	3.74E-08	4.7160	7.6027	9.1337	chr11	85405265	85522184	-	NM_001162951	SYTL2
TC05002656.hg.1	4.1130	13.9296	5.33E-12	1.37E-08	6.6027	10.6394	11.7181	chr5	139554736	139623371	+	---	---
TC21000942.hg.1	4.1034	7.7519	1.47E-07	3.14E-05	3.7456	5.4679	9.8031	chr21	36089937	36090503	-	---	---
TC04001270.hg.1	4.0581	4.8162	9.55E-05	2.22E-03	6.1255	10.7143	10.5084	chr4	71521258	71547534	-	NM_144646	IGJ
TC12003252.hg.1	4.0283	9.9172	2.52E-09	1.66E-06	3.6433	8.6170	4.7328	chr12	70219084	70352503	+	BC143553	MYRFL
TC04002595.hg.1	4.0012	6.3404	2.91E-06	2.30E-04	2.9016	6.9788	4.8625	chr4	76103904	76105665	-	---	---

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC08000721.hg.1	3.9930	6.1669	4.27E-06	2.93E-04	4.5042	5.6046	8.5598	chr8	124864227	125132302	+	NM_001039112	FER1L6
TC06000970.hg.1	3.9707	6.5406	1.88E-06	1.70E-04	3.4092	8.2051	6.1367	chr6	127840600	127912962	+	NM_001010905	C6orf58
TC13000576.hg.1	3.9688	7.0120	6.81E-07	8.76E-05	3.9294	4.5442	8.6068	chr13	38136719	38183563	-	NM_001135934	POSTN
TC04001193.hg.1	3.9674	12.7089	2.95E-11	4.07E-08	2.7636	6.9652	5.7111	chr4	52859866	52883786	-	NM_001024611	LRRC66
TC01003125.hg.1	3.9519	10.1579	1.66E-09	1.21E-06	3.7925	7.7597	7.3715	chr1	146655884	146697230	-	NM_001461	FMO5
TC01000939.hg.1	3.9465	12.8585	2.38E-11	3.74E-08	4.7383	8.5278	9.1496	chr1	109656585	109749403	+	NM_020775	KIAA1324
TC07003085.hg.1	3.9250	8.7804	1.98E-08	7.43E-06	5.3184	10.3073	6.5764	chr7	100547228	100550424	-	---	---
TC12002772.hg.1	3.9199	9.3072	7.47E-09	3.79E-06	4.6866	9.3261	4.3949	chr12	14765566	14849497	-	---	---
TC06001495.hg.1	3.9128	11.4917	1.86E-10	2.03E-07	3.9586	7.3347	6.2694	chr6	30070674	30080883	-	NM_007028	TRIM31
TC05001963.hg.1	3.9119	6.1538	4.40E-06	2.98E-04	4.2769	4.4892	4.6522	chr5	151771093	151812929	-	NM_020167	NMUR2
TC11002989.hg.1	3.9044	7.0843	5.84E-07	7.85E-05	8.0050	8.7442	12.1725	chr11	1212782	1213779	-	---	---
TC12000268.hg.1	3.8945	7.8480	1.21E-07	2.73E-05	3.7509	6.4194	8.6899	chr12	27849428	27850566	+	NM_001029874	REP15
TC02004978.hg.1	3.8910	4.8217	9.42E-05	2.19E-03	5.8691	11.4305	10.1621	chr2	89998789	89999564	+	ENST00000453166	IGKV2D-28
TC12001271.hg.1	3.8808	9.2365	8.50E-09	4.16E-06	4.5943	9.1396	4.3426	chr12	14765568	14849519	-	NM_004963	GUCY2C
TC06003665.hg.1	3.8797	6.6313	1.54E-06	1.50E-04	2.8288	8.2129	5.0389	chr6	45866380	45866814	-	---	---
TC12001909.hg.1	3.8787	14.7349	1.84E-12	5.98E-09	4.0970	8.6060	8.9983	chr12	105196331	105352522	-	NM_032148	SLC41A2
TC04002940.hg.1	3.8734	6.8853	8.91E-07	1.07E-04	4.1950	7.1497	7.2337	chr4	15969849	16086001	-	NM_006017	PROM1
TC11000788.hg.1	3.8730	9.1970	9.13E-09	4.34E-06	4.6610	8.3474	7.1957	chr11	74166393	74167870	+	---	---
TC02004946.hg.1	3.8704	4.7651	1.08E-04	2.39E-03	6.1313	11.6365	10.5231	chr2	89521179	89521942	-	ENST00000482769	IGKV2-28
Top 100 down-regulated transcript clusters from the N v BE comparison (differentiating Barrett's mucosa from normal squamous epithelium)													
TC4_ctg9_hap1000001.hg.1	-5.0713	-5.2148	3.75E-05	1.19E-03	11.0813	3.4932	3.3825	chr4_ctg9_hap1	143089	193241	+	NM_014058	TMPRSS11E
TC04000369.hg.1	-5.0586	-5.2249	3.66E-05	1.17E-03	11.0680	3.4914	3.4121	chr4	69313167	69363322	+	NM_014058	TMPRSS11E
TC04001258.hg.1	-5.0294	-4.7587	1.09E-04	2.42E-03	10.7669	2.6350	2.6667	chr4	69092371	69111438	-	NM_182502	TMPRSS11B
TC18000124.hg.1	-4.7607	-4.6051	1.57E-04	3.09E-03	10.2571	2.9679	2.7953	chr18	28898052	28937393	+	NM_001942	DSG1
TC01005834.hg.1	-4.7172	-4.2928	3.30E-04	5.11E-03	11.2587	3.5787	3.0263	chr1	153012201	153013594	-	---	---
TC01000821.hg.1	-4.7092	-5.0845	5.08E-05	1.44E-03	10.2816	3.3565	2.9857	chr1	86889769	86922241	+	NM_006536	CLCA2
TC01004511.hg.1	-4.7089	-4.7222	1.19E-04	2.57E-03	11.5816	3.6352	2.8749	chr1	87012759	87046432	+	---	---
TC18000435.hg.1	-4.6733	-5.3001	3.08E-05	1.05E-03	10.1880	3.4078	3.0548	chr18	28570052	28622781	-	NM_001941	DSC3
TC04001253.hg.1	-4.6299	-5.2538	3.42E-05	1.13E-03	9.5993	2.8341	2.7737	chr4	68775103	68829858	-	NM_001114387	TMPRSS11A
TC01000835.hg.1	-4.5937	-5.2733	3.27E-05	1.09E-03	10.5098	3.6973	3.8221	chr1	89829436	89853719	+	NM_198460	GBP6
TC02000716.hg.1	-4.5936	-5.3643	2.65E-05	9.49E-04	11.5806	4.5013	4.5807	chr2	113763038	113766126	+	BC107043	IL36A
TC12000136.hg.1	-4.5715	-5.1171	4.71E-05	1.38E-03	10.9389	3.8276	3.8357	chr12	8975068	9039597	+	NM_144670	A2ML1
TC01000823.hg.1	-4.5600	-4.7881	1.02E-04	2.31E-03	11.0913	3.5421	2.8544	chr1	87012759	87046437	+	NM_012128	CLCA4
TC01006377.hg.1	-4.5249	-4.1584	4.54E-04	6.34E-03	11.6230	3.7703	4.1079	chr1	153028589	153030013	-	NM_005988	SPRR2A
TC18000228.hg.1	-4.4808	-5.1902	3.97E-05	1.23E-03	10.1920	3.3984	3.4266	chr18	61254223	61271873	+	NM_012397	SERPINB13

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC04001252.hg.1	-4.4576	-4.9514	6.94E-05	1.78E-03	10.3312	3.9104	3.7742	chr4	68686594	68749750	-	NM_004262	TMPRSS11D
TC13000295.hg.1	-4.3241	-4.8670	8.47E-05	2.04E-03	10.3551	2.7336	3.3748	chr13	78049967	78219398	+	NM_001160706	SCEL
TC01003247.hg.1	-4.2956	-4.9118	7.62E-05	1.90E-03	11.7091	5.0342	5.0450	chr1	152381719	152386750	-	NM_016190	CRNN
TC20000023.hg.1	-4.2841	-4.7705	1.06E-04	2.37E-03	11.6566	4.9902	4.8450	chr20	2276613	2321725	+	NM_003245	TGM3
TC18001005.hg.1	-4.2761	-4.5580	1.76E-04	3.33E-03	10.3935	3.4489	3.5458	chr18	61322431	61329197	-	NM_006919	SERPINB3
TC6_mcf_hap5000038.hg.1	-4.1744	-4.9318	7.27E-05	1.84E-03	11.4233	4.8206	4.8134	chr6_mcf_hap5	2333431	2339631	+	NM_001010909	MUC21
TC16000456.hg.1	-4.1692	-5.3772	2.57E-05	9.28E-04	10.0465	3.7904	3.6352	chr16	55600584	55601599	+	NM_032330	CAPNS2
TC6_apd_hap1000031.hg.1	-4.1663	-4.9111	7.63E-05	1.90E-03	11.4397	4.8544	4.8559	chr6_apd_hap1	2263124	2269314	+	NM_001010909	MUC21
TC6_dbb_hap3000044.hg.1	-4.1663	-4.9111	7.63E-05	1.90E-03	11.4397	4.8544	4.8559	chr6_dbb_hap3	2245596	2251786	+	NM_001010909	MUC21
TC01003260.hg.1	-4.1316	-5.1570	4.29E-05	1.29E-03	9.8320	3.9570	3.8919	chr1	153346181	153348125	-	NM_005621	S100A12
TC17001468.hg.1	-4.1182	-4.2067	4.05E-04	5.88E-03	9.2495	3.7841	3.5375	chr17	38854243	38860002	-	NM_019016	KRT24
TC18000188.hg.1	-4.0953	-5.0717	5.24E-05	1.47E-03	8.8368	2.8528	2.7814	chr18	52258390	52266724	+	NM_173629	DYNAP
TC18001004.hg.1	-4.0845	-4.3780	2.70E-04	4.45E-03	9.7904	3.3744	3.3226	chr18	61304493	61311532	-	NM_002974	SERPINB4
TC06001785.hg.1	-4.0489	-4.1690	4.43E-04	6.25E-03	11.0703	3.4573	3.2183	chr6	49695092	49712168	-	NM_001190986	CRISP3
TC04001255.hg.1	-4.0078	-5.6016	1.53E-05	6.54E-04	8.0668	2.4031	2.4842	chr4	68918916	68995598	-	NM_207407	TMPRSS11F
TC01004752.hg.1	-3.9937	-3.8459	9.55E-04	1.06E-02	12.1839	3.7808	6.1204	chr1	152957973	152958289	+	---	---
TC6_cox_hap2000051.hg.1	-3.9710	-4.8046	9.81E-05	2.25E-03	11.7406	5.2996	5.5784	chr6_cox_hap2	2463385	2469590	+	NM_001010909	MUC21
TC12001523.hg.1	-3.9690	-4.4580	2.23E-04	3.89E-03	11.6105	5.0357	4.9967	chr12	52840435	52845971	-	NM_005555	KRT6B
TC12001536.hg.1	-3.9300	-4.3245	3.06E-04	4.86E-03	11.9955	5.1747	5.0701	chr12	53200327	53208335	-	NM_002272	KRT4
TC02001715.hg.1	-3.9147	-8.0748	7.73E-08	2.04E-05	8.9278	3.9288	3.8544	chr2	31395922	31456724	-	NM_001145122	CAPN14
TC04000413.hg.1	-3.9141	-4.8374	9.08E-05	2.14E-03	8.2143	2.3131	2.2276	chr4	75174190	75181024	+	NM_001013442	EPGN
TC19001446.hg.1	-3.9071	-4.8353	9.13E-05	2.15E-03	11.0953	4.9862	4.8912	chr19	36014269	36019253	-	NM_198538	SBSN
TC11001505.hg.1	-3.8991	-5.6662	1.32E-05	5.95E-04	8.8387	3.0188	2.9702	chr11	26580579	26593815	-	NM_001135091	MUC15
TC19001572.hg.1	-3.8938	-5.1939	3.94E-05	1.23E-03	12.0359	6.1708	5.9529	chr19	42891171	42894444	-	NM_032488	CNFN
TC01001244.hg.1	-3.8731	-5.6906	1.25E-05	5.72E-04	9.4470	3.8030	3.9397	chr1	152881021	152884362	+	NM_005547	IVL
TC12001525.hg.1	-3.8673	-4.5220	1.92E-04	3.52E-03	11.9040	5.6935	5.6717	chr12	52880958	52887181	-	NM_005554	KRT6A
TC01005829.hg.1	-3.8562	-3.8410	9.66E-04	1.07E-02	9.5017	3.4639	3.4949	chr1	152286184	152286725	-	---	---
TC12002888.hg.1	-3.8372	-4.7673	1.07E-04	2.38E-03	12.1279	6.0879	6.1057	chr12	52912731	52912944	-	---	---
TC6_mann_hap4000046.hg.1	-3.8331	-4.7615	1.09E-04	2.41E-03	11.7601	5.4994	5.7704	chr6_mann_hap4	2299571	2305861	+	NM_001010909	MUC21
TC6_qbl_hap6000044.hg.1	-3.7804	-4.7032	1.25E-04	2.64E-03	11.7903	5.4797	5.9121	chr6_qbl_hap6	2244470	2250865	+	NM_001010909	MUC21
TC05000805.hg.1	-3.7629	-4.6387	1.45E-04	2.92E-03	11.1755	4.3385	5.7582	chr5	147405246	147516925	+	NM_001127699	SPINK5
TC01006375.hg.1	-3.7361	-4.2882	3.34E-04	5.15E-03	10.1894	3.8505	4.2319	chr1	153042704	153044111	-	NM_001017418	SPRR2B
TC17000513.hg.1	-3.7118	-4.8641	8.53E-05	2.05E-03	11.1606	5.1270	4.9398	chr17	39657235	39658661	+	OTTHUMT000002	OTTHUMG0000

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
												57901	0133684
TC04001257.hg.1	-3.7040	-5.3486	2.75E-05	9.71E-04	7.4605	1.8600	2.0223	chr4	69054242	69083798	-	NM_001129907	TMPRSS11BNL
TC17001508.hg.1	-3.6987	-4.3286	3.03E-04	4.83E-03	12.0644	5.4486	5.5278	chr17	39657233	39661865	-	NM_002274	KRT13
TC04002578.hg.1	-3.6828	-5.0243	5.85E-05	1.59E-03	7.2778	1.7763	1.7791	chr4	69049896	69051921	-	---	---
TC15001829.hg.1	-3.6700	-5.3922	2.48E-05	9.06E-04	10.6656	5.1489	5.1443	chr15	89998680	90039844	-	NM_016321	RHCG
TC12001526.hg.1	-3.6606	-4.9345	7.23E-05	1.83E-03	11.5277	5.4670	5.3992	chr12	52908359	52914471	-	NM_000424	KRT5
TC01001488.hg.1	-3.6170	-4.8264	9.32E-05	2.18E-03	9.8172	4.2112	4.4631	chr1	171154347	171181822	+	NM_001460	FMO2
TC12001538.hg.1	-3.5802	-6.1160	4.79E-06	3.12E-04	10.7002	5.5556	5.4813	chr12	53231588	53242778	-	NM_173352	KRT78
TC01005830.hg.1	-3.5797	-3.7231	1.28E-03	1.30E-02	7.3810	2.0343	2.2763	chr1	152287087	152287969	-	---	---
TC01006378.hg.1	-3.5579	-3.8522	9.41E-04	1.05E-02	9.3667	3.3687	3.4908	chr1	153112594	153113969	-	NR_003062	SPRR2C
TC06000353.hg.1	-3.5562	-4.7864	1.02E-04	2.31E-03	11.8250	6.0283	6.1762	chr6	30951485	30957680	+	NM_001010909	MUC21
TC01001246.hg.1	-3.5268	-4.1143	5.05E-04	6.84E-03	12.0565	5.5651	6.4198	chr1	152956557	152958290	+	NM_001199828	SPRR1A
TC05000810.hg.1	-3.5150	-5.0899	5.02E-05	1.43E-03	9.3987	3.6424	4.7522	chr5	147691982	147695485	+	NM_032566	SPINK7
TC01005836.hg.1	-3.5147	-3.8589	9.26E-04	1.04E-02	9.2679	3.4210	3.6068	chr1	153112594	153113969	-	---	---
TC01001254.hg.1	-3.5096	-5.3015	3.06E-05	1.04E-03	10.9544	5.9423	5.8387	chr1	153330330	153333503	+	NM_002965	S100A9
TC01006366.hg.1	-3.5071	-6.1712	4.23E-06	2.91E-04	9.0579	4.3205	4.4973	chr1	115125469	115213043	-	NM_001256404	DENND2C
TC12001524.hg.1	-3.5012	-4.4113	2.49E-04	4.20E-03	11.7605	6.2612	6.0276	chr12	52862300	52867569	-	NM_173086	KRT6C
TC01006374.hg.1	-3.4836	-4.2407	3.74E-04	5.57E-03	9.8343	4.0621	4.0546	chr1	153012201	153014407	-	NM_006945	SPRR2D
TC0X001579.hg.1	-3.4744	-5.9209	7.42E-06	4.10E-04	8.9929	3.9831	3.7414	chrX	2670337	2693037	+	---	---
TC10002963.hg.1	-3.4630	-6.2656	3.43E-06	2.55E-04	10.0795	5.1799	5.0563	chr10	47157983	47174122	-	NM_001098845	ANXA8L1
TC10000314.hg.1	-3.4342	-6.2694	3.40E-06	2.54E-04	9.6810	4.9928	4.7928	chr10	48255225	48279199	+	NM_001098845	ANXA8L1
TC13001343.hg.1	-3.4336	-5.1450	4.41E-05	1.32E-03	9.5388	4.1725	4.2259	chr13	20797502	20806372	-	---	---
TC01003255.hg.1	-3.4272	-3.8276	9.97E-04	1.09E-02	10.3695	4.7888	4.6191	chr1	153065611	153078660	-	NM_001024209	SPRR2E
TC03001905.hg.1	-3.4075	-6.1257	4.69E-06	3.09E-04	8.2332	3.5838	3.3951	chr3	151011876	151034740	-	NM_023915	GPR87
TC01001247.hg.1	-3.4056	-4.2733	3.46E-04	5.29E-03	11.4350	5.0623	5.1009	chr1	152974223	152976332	+	NM_001097589	SPRR3
TC10000309.hg.1	-3.3714	-6.3011	3.17E-06	2.43E-04	9.6987	5.1796	4.9404	chr10	47742393	47770871	+	NM_001630	ANXA8L2
TC14000971.hg.1	-3.3575	-5.4588	2.13E-05	8.17E-04	9.9252	5.0235	4.9707	chr14	24718320	24732416	-	NM_000359	TGM1
TC19000538.hg.1	-3.3296	-5.9784	6.52E-06	3.77E-04	10.0305	5.6299	5.2216	chr19	39687601	39692524	+	NM_001001414	NCCRP1
TC06003208.hg.1	-3.3166	-5.2305	3.62E-05	1.16E-03	7.6543	2.7852	2.7239	chr6	168084231	168088891	+	---	---
TC01001248.hg.1	-3.2934	-4.3953	2.59E-04	4.32E-03	11.9041	6.4651	6.5435	chr1	153003678	153005376	+	NM_003125	SPRR1B
TC01006376.hg.1	-3.2793	-3.2693	3.70E-03	2.64E-02	9.3937	3.6838	3.8991	chr1	153084590	153085991	-	NM_001014450	SPRR2F
TC02002218.hg.1	-3.2732	-4.9750	6.57E-05	1.72E-03	7.3787	2.8417	2.7842	chr2	113531492	113542971	-	NM_000575	IL1A
TC02003832.hg.1	-3.2487	-4.9517	6.94E-05	1.78E-03	9.5578	4.9481	5.2620	chr2	218999678	219000012	+	---	---
TC06003648.hg.1	-3.2371	-6.1339	4.60E-06	3.06E-04	8.8855	4.4670	5.8041	chr6	42059976	42061997	-	---	---
TC08001637.hg.1	-3.2269	-5.3131	2.98E-05	1.02E-03	8.1732	3.4877	3.4292	chr8	130760442	130799134	-	NM_031415	GSDMC
TC11002740.hg.1	-3.2267	-5.6451	1.39E-05	6.14E-04	10.8745	6.3786	5.9918	chr11	67429963	67430223	+	---	---
TC12002789.hg.1	-3.1721	-4.2239	3.89E-04	5.72E-03	7.5141	2.8223	2.7429	chr12	21689123	21757554	-	---	---
TC01004997.hg.1	-3.1476	-5.1754	4.11E-05	1.26E-03	8.7065	4.2472	4.1867	chr1	209602165	209606183	+	---	---

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC01005828.hg.1	-3.1353	-3.8080	1.04E-03	1.13E-02	9.6127	4.8973	4.6607	chr1	152274640	152281460	-	---	---
TC01001199.hg.1	-3.1038	-6.2147	3.84E-06	2.73E-04	8.4876	4.1976	4.1870	chr1	151009029	151020076	+	NM_001159642	BNIPL
TC12001436.hg.1	-3.0994	-6.2087	3.89E-06	2.75E-04	8.0995	3.9126	3.8091	chr12	48103517	48119355	-	NM_001172439	ENDOU
TC18000127.hg.1	-3.0992	-4.6099	1.55E-04	3.06E-03	11.2379	6.7588	3.1241	chr18	29027732	29058665	+	NM_001944	DSG3
TC03001042.hg.1	-3.0927	-5.5112	1.89E-05	7.52E-04	8.7630	4.3226	4.2599	chr3	189349205	189615068	+	NM_003722	TP63
TC04000282.hg.1	-3.0808	-5.4026	2.42E-05	8.91E-04	8.5796	4.4648	6.5740	chr4	48988264	49064098	+	NM_025087	CWH43
TC12002078.hg.1	-3.0776	-4.8061	9.78E-05	2.24E-03	8.2826	4.1435	4.1012	chr12	123185840	123187904	-	NM_177551	HCAR2
TC06004001.hg.1	-3.0511	-5.2619	3.36E-05	1.11E-03	8.4543	3.8965	3.6994	chr6	168080306	168096970	-	---	---
TC12001297.hg.1	-3.0497	-4.1367	4.78E-04	6.59E-03	7.0979	2.6419	2.5754	chr12	21689123	21757781	-	NM_021957	GYS2
TC06003638.hg.1	-3.0310	-6.1348	4.59E-06	3.06E-04	6.5756	2.3900	2.4994	chr6	40825508	40827706	-	---	---
TC02000963.hg.1	-3.0084	-4.9519	6.94E-05	1.78E-03	9.5383	5.3231	5.1732	chr2	159651829	159719293	+	NM_001017920	DAPL1
TC15001996.hg.1	-3.0011	-6.3750	2.70E-06	2.19E-04	7.8457	3.7938	3.6918	chr15	100940600	101084925	-	NM_178842	CERS3
TC12000189.hg.1	-2.9943	-5.3439	2.78E-05	9.77E-04	10.7819	6.5816	5.7443	chr12	13349602	13369708	+	NM_001423	EMP1
TC04002094.hg.1	-2.9936	-5.2595	3.38E-05	1.11E-03	9.6699	5.2431	6.6181	chr4	87515885	87736302	+	---	---
TC6_mann_hap4000176.hg.1	-2.9905	-4.5911	1.63E-04	3.16E-03	6.8706	3.0106	2.5619	chr6_mann_hap4	2370339	2376003	+	---	---

Table 3: Top 100 up- and down-regulated transcript clusters for differentiating intervention-requiring disease from non-dysplastic Barrett's

553 up-regulated and 297 down-regulated transcript clusters were identified as differentially expressed between non-dysplastic Barrett's mucosa and intervention requiring disease (BE v HGD-EAC, adjusted p-value < 0.05 and logfc > log2(1.5). Tx: transcript, logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, Exp_{base}: average expression in baseline samples, here, non-dysplastic Barrett's mucosa, Exp_{duo}: average expression in duodenal epithelium, Exp_{stom}: average expression in proximal stomach epithelium, str: strand. Top 100 up- and down-regulated transcript clusters based on greatest |logfc|; listed in order of decreasing |logfc|.

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
Top 100 up-regulated transcript clusters from the BE v HGD-EAC comparison (differentiating intervention requiring disease from non-dysplastic Barrett's mucosa)													
TC10000008.hg.1	0.9875	9.1616	9.75E-09	9.40E-05	6.7622	6.4083	6.8725	chr10	1034338	1065876	+	NM_012341	GTPBP4
TC11002233.hg.1	3.1433	8.6302	2.63E-08	1.78E-04	3.2374	3.7780	4.7078	chr11	102641233	102651359	-	NM_002425	MMP10
TC20001062.hg.1	0.8791	8.5569	3.03E-08	1.86E-04	7.7647	7.4875	7.6931	chr20	2633178	2639039	+	---	---
TC11002234.hg.1	3.8424	8.4301	3.86E-08	1.86E-04	3.8678	5.8857	5.4730	chr11	102660641	102668966	-	NM_001145938	MMP1
TC09000691.hg.1	0.6903	8.4675	3.59E-08	1.86E-04	6.4568	6.2320	6.4354	chr9	131217434	131263571	+	NM_001242352	ODF2
TC07001654.hg.1	0.8182	8.3177	4.80E-08	2.03E-04	6.3423	6.3634	6.2829	chr7	99690351	99699563	-	NM_182776	MCM7
TC02001579.hg.1	0.8089	8.3384	4.61E-08	2.03E-04	4.2422	4.2419	4.4507	chr2	11584501	11606297	-	NM_198256	E2F6
TC20000637.hg.1	0.7846	7.7811	1.39E-07	3.52E-04	7.8419	7.8910	7.9649	chr20	13694969	13765565	-	NM_016649	ESF1
TC03003098.hg.1	0.7644	7.7738	1.41E-07	3.52E-04	7.4047	7.2070	7.3635	chr3	133319450	133380690	-	---	---

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC03001796.hg.1	0.6654	7.8514	1.21E-07	3.52E-04	6.8308	6.6603	6.7833	chr3	133317019	133380737	-	NM_007027	TOPBP1
TC20000026.hg.1	0.6631	7.8418	1.23E-07	3.52E-04	6.3135	6.2026	6.2336	chr20	2632791	2639039	+	ENST00000413522	SNORD56
TC01003200.hg.1	0.7385	7.7480	1.48E-07	3.58E-04	8.3374	8.1723	8.0951	chr1	150190717	150208504	-	NM_030920	ANP32E
TC03001632.hg.1	1.3822	7.7119	1.60E-07	3.72E-04	4.6369	4.2227	4.4731	chr3	108268718	108308491	-	NM_020890	KIAA1524
TC03000930.hg.1	1.2695	7.4383	2.79E-07	5.42E-04	6.8198	5.8232	6.3298	chr3	172468472	172539264	+	NM_018098	ECT2
TC20000063.hg.1	1.0923	7.4438	2.76E-07	5.42E-04	5.5652	5.7772	5.7838	chr20	5931298	5975852	+	NM_032485	MCM8
TC07002911.hg.1	3.5651	7.4207	2.90E-07	5.43E-04	3.8273	5.5731	4.3933	chr7	41724713	41726643	-	---	---
TC19001301.hg.1	0.6634	7.3541	3.32E-07	5.91E-04	6.2860	6.2293	6.5576	chr19	19312224	19314238	-	NM_176880	NR2C2AP
TC09002753.hg.1	0.7617	7.1391	5.21E-07	8.58E-04	5.9991	6.0231	5.9738	chr9	116169518	116172955	-	---	---
TC02001139.hg.1	0.9127	7.0213	6.68E-07	1.02E-03	5.8863	5.7925	5.5752	chr2	196440701	196602426	+	NM_020342	SLC39A10
TC16000634.hg.1	0.9658	6.9882	7.16E-07	1.07E-03	6.5930	6.2274	6.5186	chr16	81040103	81066719	+	NM_001100624	CENPN
TC01002529.hg.1	0.6280	6.8853	8.91E-07	1.28E-03	6.7942	6.6781	6.8074	chr1	38422647	38456593	-	NM_006802	SF3A3
TC11002237.hg.1	2.8166	6.8179	1.03E-06	1.42E-03	5.6483	4.1360	3.7047	chr11	102733464	102745764	-	NM_002426	MMP12
TC08001507.hg.1	0.7486	6.8041	1.06E-06	1.43E-03	6.0614	6.1313	6.1001	chr8	104410863	104427468	-	NM_030780	SLC25A32
TC11003311.hg.1	3.0850	6.7924	1.09E-06	1.44E-03	5.0643	6.3095	5.9152	chr11	102667775	102668070	-	---	---
TC08001584.hg.1	1.3620	6.7240	1.26E-06	1.64E-03	6.0454	5.7622	6.3661	chr8	124332090	124428590	-	NM_014109	ATAD2
TC06003842.hg.1	1.2737	6.7106	1.30E-06	1.65E-03	2.8942	4.0897	3.3473	chr6	126497759	126498481	-	---	---
TC01003679.hg.1	1.4503	6.6842	1.37E-06	1.72E-03	4.8733	4.2871	4.3622	chr1	200520625	200589862	-	NM_014875	KIF14
TC09001632.hg.1	0.6940	6.6691	1.42E-06	1.74E-03	6.9215	6.8208	6.9426	chr9	131231236	131235291	-	OTTHUMT00000054431	OTTHUMG00000020745
TC01003416.hg.1	0.7756	6.6586	1.45E-06	1.75E-03	7.8658	7.7989	7.9958	chr1	161070346	161087901	-	NM_012394	PFDN2
TC09000573.hg.1	0.8191	6.6427	1.50E-06	1.78E-03	6.5976	6.3365	6.4034	chr9	116037623	116055466	+	NM_001244926	PRPF4
TC11003312.hg.1	3.0525	6.5873	1.69E-06	1.85E-03	3.5618	4.9599	4.7804	chr11	102668128	102668877	-	---	---
TC01002500.hg.1	1.2475	6.5860	1.70E-06	1.85E-03	5.6019	5.3196	5.0381	chr1	36185819	36235568	-	NM_001190481	CLSPN
TC08000701.hg.1	1.1968	6.6028	1.64E-06	1.85E-03	4.3672	4.2573	4.6509	chr8	121457640	121554373	+	NM_022045	MTBP
TC10000729.hg.1	0.6551	6.5887	1.69E-06	1.85E-03	5.2822	5.3109	5.3055	chr10	102747124	102754158	+	NM_001163812	C10orf2
TC03000503.hg.1	0.7602	6.5738	1.74E-06	1.87E-03	4.9040	4.8526	5.2280	chr3	99536678	99897476	+	NM_001167924	CMSS1
TC12001106.hg.1	1.1225	6.5353	1.90E-06	1.97E-03	6.0030	5.7554	5.6722	chr12	2966847	2986321	-	NM_001243088	FOXM1
TC0X000921.hg.1	0.9784	6.4914	2.09E-06	2.14E-03	5.4755	5.8402	6.1915	chrX	23720370	23784592	-	NM_001033583	ACOT9
TC03001751.hg.1	0.6980	6.4765	2.16E-06	2.17E-03	5.9481	5.5719	5.7402	chr3	127783625	127872757	-	NM_003707	RUVBL1
TC02002445.hg.1	0.8094	6.4485	2.29E-06	2.23E-03	5.3803	5.3321	5.4994	chr2	157180944	157198860	-	NM_006186	NR4A2
TC12002747.hg.1	0.7529	6.4366	2.35E-06	2.24E-03	4.0666	4.2865	3.9532	chr12	9546891	9566384	-	---	---
TC17001291.hg.1	1.1905	6.3996	2.55E-06	2.25E-03	3.5389	3.7734	3.1448	chr17	26598641	26598673	-	---	---
TC20000357.hg.1	0.8815	6.4182	2.45E-06	2.25E-03	6.3500	6.3416	6.1610	chr20	44441215	44445596	+	NM_007019	UBE2C
TC12001142.hg.1	0.7720	6.4056	2.52E-06	2.25E-03	6.1129	5.8878	5.9953	chr12	6666029	6677857	-	NM_001033714	NOP2
TC20000385.hg.1	0.6363	6.4276	2.40E-06	2.25E-03	6.6297	6.6035	6.5681	chr20	47835832	47860614	+	NM_017895	DDX27
TC19002224.hg.1	0.8162	6.3685	2.73E-06	2.33E-03	6.6218	6.7176	7.1703	chr19	53935227	53947925	+	---	---

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC10000900.hg.1	0.6798	6.3650	2.76E-06	2.33E-03	7.2043	6.9398	7.0861	chr10	124913760	124924886	+	NM_001007793	BUB3
TC20000198.hg.1	1.3386	6.3085	3.12E-06	2.49E-03	7.1837	6.4453	6.2455	chr20	30326904	30389608	+	NM_012112	TPX2
TC08001528.hg.1	1.1368	6.2984	3.19E-06	2.49E-03	5.5913	5.5809	5.9407	chr8	110253148	110346614	-	NM_001128211	NUDCD1
TC09002191.hg.1	0.8368	6.2955	3.21E-06	2.49E-03	8.0579	7.8427	8.0115	chr9	116170093	116170786	+	---	---
TC07001177.hg.1	0.8028	6.3257	3.01E-06	2.49E-03	5.4768	5.3166	5.6015	chr7	19735085	19748710	-	NM_001002926	TWISTNB
TC08002008.hg.1	0.7956	6.2507	3.55E-06	2.63E-03	7.8715	7.2982	7.6158	chr8	90770370	90782052	+	---	---
TC12002211.hg.1	0.7756	6.2576	3.49E-06	2.63E-03	4.6182	5.0041	5.0178	chr12	7087683	7089096	+	---	---
TC02004051.hg.1	0.6975	6.2550	3.51E-06	2.63E-03	5.1491	5.2019	5.3123	chr2	11584501	11606297	-	---	---
TC19000443.hg.1	0.6860	6.2049	3.93E-06	2.85E-03	6.3569	6.0960	6.5430	chr19	34895303	34917072	+	NM_032346	PDCD2L
TC15001350.hg.1	0.6352	6.1983	3.99E-06	2.86E-03	6.5864	6.5122	6.6898	chr15	50569389	50647605	-	NM_016654	GABPB1
TC07002137.hg.1	1.4656	6.1740	4.21E-06	2.99E-03	3.6406	3.5168	3.4445	chr7	8301863	8384146	+	---	---
TC03000518.hg.1	0.6228	6.1261	4.68E-06	3.26E-03	6.5662	6.6896	6.8116	chr3	101280706	101285290	+	NM_017819	TRMT10C
TC01001430.hg.1	1.4352	6.1116	4.84E-06	3.33E-03	5.7653	5.1832	5.1570	chr1	163291690	163325554	+	NM_031423	NUF2
TC07001259.hg.1	0.8257	6.0861	5.12E-06	3.47E-03	5.3532	5.3228	5.4833	chr7	32620553	32758780	-	NR_036680	DPY19L1P1
TC17001346.hg.1	0.6022	6.0849	5.13E-06	3.47E-03	6.2284	6.1506	6.2170	chr17	30187923	30228731	-	NM_018428	UTP6
TC05000154.hg.1	0.6040	6.0772	5.22E-06	3.49E-03	6.9486	6.7317	6.9989	chr5	34915481	34926101	+	NM_018321	BRIX1
TC03000114.hg.1	1.0977	6.0570	5.46E-06	3.58E-03	6.0911	6.0151	5.9534	chr3	20215736	20227919	+	ENST00000441442	SGOL1-AS1
TC11002235.hg.1	3.4700	6.0013	6.19E-06	3.77E-03	3.3712	4.3371	3.9633	chr11	102706528	102714534	-	NM_002422	MMP3
TC08001560.hg.1	1.1185	6.0205	5.93E-06	3.77E-03	3.8235	3.6226	4.0013	chr8	120846181	120868250	-	NM_024094	DSCC1
TC20001440.hg.1	0.9493	6.0044	6.15E-06	3.77E-03	6.5798	6.9932	6.7147	chr20	13763456	13765569	-	---	---
TC03000242.hg.1	1.1619	5.9463	7.01E-06	3.99E-03	5.4515	5.0005	4.9262	chr3	44803209	44914868	+	NM_020242	KIF15
TC05001352.hg.1	0.9225	5.9475	6.99E-06	3.99E-03	2.7944	2.9962	3.1393	chr5	54273692	54281491	-	NM_001135604	ESM1
TC01003798.hg.1	1.3213	5.9368	7.16E-06	4.00E-03	5.2409	5.0854	4.9558	chr1	211831599	211848972	-	NM_001204182	NEK2
TC07000384.hg.1	0.9517	5.9121	7.57E-06	4.02E-03	5.6995	6.0298	5.9177	chr7	64126501	64171404	+	NM_001013746	ZNF107
TC06001799.hg.1	0.7835	5.9106	7.60E-06	4.02E-03	6.6114	6.4119	6.5477	chr6	52128807	52149582	-	NM_002388	MCM3
TC15000875.hg.1	0.7109	5.9099	7.61E-06	4.02E-03	5.5139	5.5235	5.4312	chr15	91509575	91531854	+	ENST00000554388	PRC1-AS1
TC12001361.hg.1	1.2737	5.8587	8.54E-06	4.12E-03	7.3470	7.0351	6.8868	chr12	31433518	31479306	-	NM_001135811	FAM60A
TC01000563.hg.1	0.8931	5.8444	8.82E-06	4.22E-03	4.7675	4.5683	4.6364	chr1	45205490	45233439	+	NM_006845	KIF2C
TC04000411.hg.1	2.0652	5.8194	9.34E-06	4.32E-03	5.7771	5.7266	5.0013	chr4	74735109	74737019	+	NM_001511	CXCL1
TC01006326.hg.1	0.6552	5.8207	9.31E-06	4.32E-03	6.7922	6.8511	7.0289	chr1	149858490	149858961	+	NM_003517	HIST2H2AC
TC05001319.hg.1	1.1994	5.8063	9.62E-06	4.36E-03	3.4597	3.1126	3.4485	chr5	43486803	43515273	-	NM_198566	C5orf34
TC09001377.hg.1	1.0513	5.8113	9.51E-06	4.36E-03	5.3577	5.5816	5.7663	chr9	99401859	99417599	-	NM_153698	AAED1
TC07000159.hg.1	1.1480	5.7813	1.02E-05	4.52E-03	6.7238	6.7284	6.8919	chr7	26191847	26226756	+	NM_004289	NFE2L3
TC13000493.hg.1	0.8181	5.7688	1.05E-05	4.56E-03	5.6762	6.2843	5.5456	chr13	25456412	25497085	-	NM_018451	CENPJ
TC13000119.hg.1	1.1800	5.7610	1.07E-05	4.59E-03	6.0656	6.1441	5.9873	chr13	34392186	34540695	+	NM_181558	RFC3
TC04000408.hg.1	3.5757	5.7567	1.08E-05	4.59E-03	3.5790	3.5081	3.2133	chr4	74606223	74609433	+	NM_000584	IL8

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC05003257.hg.1	1.4744	5.7478	1.10E-05	4.62E-03	4.1786	4.5060	3.9078	chr5	137627772	137666776	-	---	---
TC03002088.hg.1	0.9307	5.7343	1.13E-05	4.66E-03	5.7318	5.9236	6.0598	chr3	186507669	186524847	-	NM_002916	RFC4
TC02000669.hg.1	0.8840	5.7230	1.16E-05	4.73E-03	3.8468	3.7762	3.9905	chr2	109403213	109493034	+	NM_144978	CCDC138
TC19001804.hg.1	0.7853	5.7229	1.16E-05	4.73E-03	6.6097	7.0707	7.0332	chr19	53268747	53290034	-	NM_198457	ZNF600
TC01001976.hg.1	1.2822	5.6891	1.26E-05	4.79E-03	5.0936	4.9226	4.5555	chr1	242011269	242058450	+	NM_003686	EXO1
TC01000461.hg.1	1.1655	5.6597	1.34E-05	4.79E-03	4.7042	4.7057	4.1676	chr1	36204990	36209177	+	OTTHUMT00000023222	OTTHUMG00000008426
TC20000871.hg.1	1.0661	5.6878	1.26E-05	4.79E-03	6.6591	6.2196	6.7346	chr20	43570771	43589127	-	NM_006809	TOMM34
TC01001476.hg.1	0.9914	5.6860	1.26E-05	4.79E-03	4.9703	4.9438	5.0705	chr1	169631245	169823221	+	NM_018186	C1orf112
TC10000664.hg.1	0.9893	5.6911	1.25E-05	4.79E-03	5.3032	4.6644	4.7940	chr10	95256369	95288849	+	NM_018131	CEP55
TC08000546.hg.1	0.9340	5.6675	1.32E-05	4.79E-03	5.9138	6.0435	6.0320	chr8	90914087	90940116	+	NM_001126111	OSGIN2
TC13001100.hg.1	0.8788	5.6584	1.35E-05	4.79E-03	7.5982	7.0408	6.5181	chr13	53029617	53035443	+	---	---
TC14001561.hg.1	0.8735	5.6717	1.31E-05	4.79E-03	6.0422	7.1370	6.5288	chr14	105324129	105325618	-	---	---
TC10000413.hg.1	0.8727	5.6702	1.31E-05	4.79E-03	8.0091	7.4114	7.8545	chr10	70715884	70744829	+	NM_004728	DDX21
TC07000771.hg.1	0.8587	5.6557	1.36E-05	4.79E-03	5.6978	6.0956	6.9274	chr7	128032601	128032635	+	---	---
TC0X000185.hg.1	0.8587	5.6557	1.36E-05	4.79E-03	5.6978	6.0956	6.9274	chrX	40217980	40218014	+	---	---
TC10000534.hg.1	0.8587	5.6557	1.36E-05	4.79E-03	5.6978	6.0956	6.9274	chr10	79539622	79539656	+	---	---
TC16000856.hg.1	0.8587	5.6557	1.36E-05	4.79E-03	5.6978	6.0956	6.9274	chr16	10207165	10207199	-	---	---
TC13000223.hg.1	0.8249	5.7019	1.22E-05	4.79E-03	7.0093	6.5832	5.9840	chr13	53029495	53050763	+	NM_001098525	CKAP2
TC01005857.hg.1	0.6864	5.6521	1.37E-05	4.80E-03	8.9411	8.7448	8.8511	chr1	156278752	156308206	-	---	---
Top 100 down-regulated transcript clusters from the BE v HGD-EAC comparison (differentiating intervention requiring disease from non-dysplastic Barrett's mucosa)													
TC05002693.hg.1	-2.8627	-13.2203	1.42E-11	9.59E-07	8.3185	7.0106	7.2005	chr5	150400124	150408543	+	---	---
TC05000837.hg.1	-1.6592	-12.0845	7.46E-11	2.52E-06	6.8565	5.9994	6.1108	chr5	150399999	150408554	+	NM_002084	GPX3
TC07001746.hg.1	-5.1100	-11.2244	2.84E-10	4.80E-06	9.7787	11.3912	5.1625	chr7	107405912	107443678	-	NM_000111	SLC26A3
TC03000208.hg.1	-1.6600	-9.7721	3.25E-09	4.39E-05	6.8234	5.8305	5.9881	chr3	40428647	40470110	+	NM_001248	ENTPD3
TC14000908.hg.1	-0.9062	-9.4068	6.24E-09	7.02E-05	7.4939	7.0233	7.1822	chr14	21484922	21539031	-	NM_201535	NDRG2
TC21000305.hg.1	-3.4770	-8.9969	1.32E-08	1.11E-04	6.3180	10.5951	3.6926	chr21	19641433	19858197	-	NM_002772	TMPRSS15
TC02000571.hg.1	-0.6011	-8.5024	3.36E-08	1.86E-04	6.8137	6.1809	6.5520	chr2	95940201	95957056	+	NM_001165977	PROM2
TC05003085.hg.1	-1.5308	-8.2195	5.81E-08	2.15E-04	4.8530	6.1163	6.3207	chr5	74337140	74341142	-	---	---
TC06001229.hg.1	-0.6272	-8.1994	6.05E-08	2.15E-04	6.7756	6.5017	6.5614	chr6	3269196	3457256	-	NM_015482	SLC22A23
TC14000076.hg.1	-1.0892	-8.1275	6.96E-08	2.35E-04	7.9139	7.4242	7.6918	chr14	21484931	21485921	+	---	---
TC06003372.hg.1	-1.0659	-8.0610	7.94E-08	2.55E-04	7.7696	7.4614	7.4056	chr6	3269986	3274708	-	---	---
TC15001837.hg.1	-2.7297	-7.8068	1.32E-07	3.52E-04	10.5301	11.3706	5.8109	chr15	90328120	90358094	-	NM_001150	ANPEP
TC02002365.hg.1	-1.1825	-7.4688	2.62E-07	5.42E-04	5.5617	5.4934	5.1223	chr2	133429372	134326034	-	NM_207363	NCKAP5
TC0X000177.hg.1	-3.7021	-7.3845	3.12E-07	5.70E-04	8.2432	9.6800	4.6208	chrX	38211736	38280703	+	NM_000531	OTC
TC03001820.hg.1	-2.1758	-7.2173	4.42E-07	7.65E-04	5.3846	6.2051	6.1396	chr3	137842560	137851229	-	NM_016161	A4GNT
TC15001318.hg.1	-1.2592	-7.1555	5.03E-07	8.49E-04	7.1036	7.3251	6.1126	chr15	45771809	45815005	-	NM_013309	SLC30A4
TC18000135.hg.1	-3.6422	-7.0897	5.78E-07	9.07E-04	7.9125	10.7710	3.6477	chr18	29769987	29800366	+	NM_005925	MEP1B

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC18000557.hg.1	-0.7296	-6.6256	1.56E-06	1.81E-03	10.0556	9.4675	9.4529	chr18	61056423	61089752	-	NM_004869	VPS4B
TC15000354.hg.1	-2.8312	-6.4667	2.20E-06	2.19E-03	8.0859	9.0364	6.5693	chr15	45544428	45568149	+	NM_004212	SLC28A2
TC05002527.hg.1	-0.7051	-6.3978	2.56E-06	2.25E-03	10.6606	10.1333	9.9384	chr5	96079244	96109116	+	---	---
TC17000834.hg.1	-0.5875	-6.4028	2.54E-06	2.25E-03	7.9664	7.8964	7.6920	chr17	72744752	72765499	+	NM_004252	SLC9A3R1
TC01000822.hg.1	-3.8973	-6.3683	2.74E-06	2.33E-03	8.3085	10.1797	3.4495	chr1	86934051	86965974	+	NM_001285	CLCA1
TC06001825.hg.1	-2.8403	-6.3126	3.09E-06	2.49E-03	6.4383	7.7389	5.6694	chr6	55618443	55740375	-	NM_021073	BMP5
TC04002642.hg.1	-3.5757	-6.2047	3.93E-06	2.85E-03	8.7208	4.0165	8.4066	chr4	100340208	100341836	-	---	---
TC14000951.hg.1	-1.2205	-6.0119	6.04E-06	3.77E-03	8.1053	8.4408	8.7194	chr14	23594504	23652850	-	NM_012244	SLC7A8
TC15001730.hg.1	-0.7402	-6.0254	5.87E-06	3.77E-03	6.6622	6.3791	5.8327	chr15	81601394	81616524	-	NM_181900	STARD5
TC14001302.hg.1	-1.6045	-5.9950	6.28E-06	3.79E-03	8.2383	8.1536	7.6704	chr14	74424713	74486102	-	NM_001249	ENTPD5
TC08000508.hg.1	-1.7874	-5.9843	6.43E-06	3.84E-03	6.4689	7.8497	6.3283	chr8	80523049	80578410	+	NM_001199214	STMN2
TC02004093.hg.1	-3.7632	-5.9802	6.49E-06	3.85E-03	6.6836	10.9422	4.8379	chr2	21225293	21225747	-	---	---
TC21001061.hg.1	-0.7517	-5.9670	6.69E-06	3.93E-03	6.2901	5.7960	6.2397	chr21	37529080	37666572	+	NM_005128	DOPEY2
TC04001412.hg.1	-3.1675	-5.9447	7.03E-06	3.99E-03	7.0260	3.4076	6.6783	chr4	100333418	100356894	-	NM_000673	ADH7
TC10001637.hg.1	-0.8103	-5.9525	6.91E-06	3.99E-03	6.5511	5.7781	5.8481	chr10	105791044	105845760	-	NR_030760	MIR936
TC04001406.hg.1	-3.4899	-5.9220	7.40E-06	4.02E-03	7.9762	9.5860	3.7062	chr4	100044808	100078949	-	NM_000670	ADH4
TC04002840.hg.1	-2.4671	-5.9088	7.63E-06	4.02E-03	10.5312	9.7103	9.9395	chr4	175411328	175444044	-	---	---
TC01000999.hg.1	-0.7786	-5.9109	7.59E-06	4.02E-03	7.2323	6.7148	7.4365	chr1	113933371	114228545	+	NM_001142782	MAGI3
TC05002526.hg.1	-1.0382	-5.8884	7.99E-06	4.06E-03	9.0568	8.2406	8.0493	chr5	96071868	96077284	+	---	---
TC17001830.hg.1	-1.3366	-5.8706	8.31E-06	4.07E-03	6.7722	7.1157	6.9622	chr17	67240576	67323323	-	NM_018672	ABCA5
TC14002346.hg.1	-0.6422	-5.8716	8.29E-06	4.07E-03	5.6582	5.3406	5.4958	chr14	74399213	74417117	-	NM_152445	FAM161B
TC12000495.hg.1	-0.9646	-5.8607	8.50E-06	4.12E-03	7.6113	7.5522	6.0394	chr12	56324756	56347811	+	NM_001345	DGKA
TC04001409.hg.1	-2.5474	-5.8421	8.87E-06	4.22E-03	6.7199	7.0291	5.5124	chr4	100197523	100212185	-	NM_000667	ADH1A
TC11002370.hg.1	-0.9567	-5.8194	9.34E-06	4.32E-03	5.9010	7.2855	5.1221	chr11	119225925	119252436	-	NM_001243759	USP2
TC01003216.hg.1	-0.5971	-5.8245	9.23E-06	4.32E-03	6.7503	6.3920	6.8309	chr1	150965682	150980854	-	NM_001040217	FAM63A
TC02001628.hg.1	-1.6405	-5.8066	9.61E-06	4.36E-03	5.2961	10.2699	4.1713	chr2	21224301	21266945	-	NM_000384	APOB
TC15002157.hg.1	-0.6898	-5.7929	9.91E-06	4.43E-03	3.6160	3.5079	3.5149	chr15	40666600	40668192	+	---	---
TC15001312.hg.1	-0.9327	-5.7747	1.03E-05	4.56E-03	3.9874	4.6933	3.8862	chr15	45543884	45571449	-	---	---
TC07001825.hg.1	-0.9444	-5.7701	1.04E-05	4.56E-03	5.8357	6.0240	5.5991	chr7	123670970	123673523	-	NM_001136002	TMEM229A
TC02001115.hg.1	-0.5906	-5.7438	1.11E-05	4.62E-03	6.0381	6.2863	5.5535	chr2	190744335	191068210	+	NM_001042519	C2orf88
TC17001017.hg.1	-0.6682	-5.7350	1.13E-05	4.66E-03	6.7733	6.7112	6.3865	chr17	3907739	4046314	-	NM_015113	ZZEF1
TC06000970.hg.1	-3.7826	-5.6878	1.26E-05	4.79E-03	7.3800	8.2051	6.1367	chr6	127840600	127912962	+	NM_001010905	C6orf58
TC04001753.hg.1	-2.1907	-5.6597	1.34E-05	4.79E-03	7.5396	6.9543	7.1422	chr4	175411328	175444305	-	NM_000860	HPGD
TC04001193.hg.1	-1.9474	-5.6945	1.24E-05	4.79E-03	6.7310	6.9652	5.7111	chr4	52859866	52883786	-	NM_001024611	LRRC66
TC10000986.hg.1	-1.6933	-5.6566	1.35E-05	4.79E-03	7.5633	6.9335	7.6024	chr10	5029967	5060225	-	NM_001135241	AKR1C2
TC17000061.hg.1	-1.1910	-5.6810	1.28E-05	4.79E-03	5.8012	6.5332	6.2881	chr17	4981754	4999669	+	NM_153018	ZFP3
TC05000498.hg.1	-0.9089	-5.6505	1.37E-05	4.80E-03	7.4607	6.7025	7.0093	chr5	102455853	102538937	+	NM_015216	PPIP5K2
TC10001315.hg.1	-1.7222	-5.6362	1.42E-05	4.91E-03	7.0503	5.1631	5.4130	chr10	61410522	61495760	-	NM_194298	SLC16A9

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC10000370.hg.1	-1.4374	-5.6304	1.44E-05	4.95E-03	6.7611	6.9069	5.7868	chr10	61815632	61820501	+	OTTHUMT00000048179	OTTHUMG00000018286
TC11002446.hg.1	-0.8382	-5.6103	1.50E-05	5.10E-03	7.4657	7.0380	6.4637	chr11	128834955	129062093	-	NM_001142685	ARHGAP32
TC05000469.hg.1	-0.6460	-5.5774	1.62E-05	5.26E-03	8.7009	8.2945	8.0797	chr5	95865525	96115299	+	NM_001042440	CAST
TC05001487.hg.1	-2.5488	-5.5520	1.72E-05	5.45E-03	8.0888	9.4578	9.3542	chr5	74321171	74326724	-	NM_016591	GCNT4
TC11002988.hg.1	-1.8063	-5.5362	1.78E-05	5.50E-03	8.0619	7.2306	5.2448	chr11	1093317	1094883	-	---	---
TC19002693.hg.1	-1.4130	-5.5275	1.82E-05	5.51E-03	7.3304	8.3278	6.3001	chr19	19254774	19384074	-	NM_001001524	TM6SF2
TC0X001696.hg.1	-0.8470	-5.5016	1.93E-05	5.77E-03	6.4186	7.7685	6.0413	chrX	48306415	48307284	+	---	---
TC05001002.hg.1	-1.2953	-5.4716	2.07E-05	5.94E-03	7.7134	8.7046	6.1478	chr5	175969512	176022975	+	NM_001171976	CDHR2
TC12000807.hg.1	-0.7025	-5.4613	2.12E-05	6.03E-03	7.4189	7.2975	7.5518	chr12	104458235	104500304	+	NM_013320	HCFC2
TC06002965.hg.1	-1.0848	-5.4506	2.17E-05	6.09E-03	6.5874	6.2996	5.3019	chr6	106808683	106969074	+	---	---
TC04001807.hg.1	-0.6537	-5.4495	2.18E-05	6.09E-03	8.6750	8.3440	7.7171	chr4	185548850	185570663	-	NM_004346	CASP3
TC08001462.hg.1	-0.8186	-5.4374	2.24E-05	6.17E-03	7.4714	6.3916	7.1347	chr8	99202061	99306621	-	NM_024759	NIPAL2
TC15002529.hg.1	-0.7137	-5.4216	2.32E-05	6.29E-03	8.1774	7.6382	8.5358	chr15	49280967	49338630	-	---	---
TC19000404.hg.1	-0.8036	-5.4144	2.36E-05	6.34E-03	6.6670	6.3440	7.2068	chr19	23945746	24014998	+	NR_003662	RPSAP58
TC10001319.hg.1	-0.8431	-5.4056	2.41E-05	6.38E-03	6.8769	7.0892	6.2729	chr10	61786056	62493284	-	NM_001149	ANK3
TC02001977.hg.1	-0.9373	-5.3896	2.50E-05	6.50E-03	6.7224	6.2162	6.3029	chr2	72403113	73053177	-	NM_015189	EXOC6B
TC17001890.hg.1	-0.7272	-5.3825	2.54E-05	6.50E-03	8.9104	9.4410	8.7395	chr17	73937589	73975515	-	NM_001185039	ACOX1
TC15001339.hg.1	-0.6384	-5.3885	2.51E-05	6.50E-03	8.0660	7.5612	8.4501	chr15	49280835	49338760	-	NM_001193489	SECISBP2L
TC02002194.hg.1	-1.4166	-5.3592	2.68E-05	6.51E-03	10.1642	9.8855	7.1844	chr2	110841447	110874143	-	NM_005434	MALL
TC06003373.hg.1	-0.6944	-5.3565	2.70E-05	6.51E-03	4.9371	4.8688	5.0064	chr6	3308907	3313510	-	---	---
TC02002487.hg.1	-0.6035	-5.3279	2.88E-05	6.75E-03	6.0139	5.7755	6.7461	chr2	165510134	165700189	-	NM_014900	COBLL1
TC21000730.hg.1	-2.2084	-5.3233	2.91E-05	6.79E-03	8.0516	8.0492	5.1748	chr21	41014313	41031839	+	---	---
TC02001000.hg.1	-1.9922	-5.3209	2.93E-05	6.80E-03	6.5570	3.4053	3.9284	chr2	169921299	169952677	+	NM_005771	DHRS9
TC01003477.hg.1	-1.5565	-5.2872	3.17E-05	7.04E-03	7.7319	8.2822	5.1185	chr1	167022073	167059868	-	NM_005814	GPA33
TC05001379.hg.1	-0.6035	-5.2596	3.38E-05	7.26E-03	6.3610	6.6917	5.6896	chr5	56215429	56267502	-	NM_152622	MIER3
TC14002093.hg.1	-1.3424	-5.2449	3.50E-05	7.40E-03	5.6459	6.7498	4.0190	chr14	76041231	76045931	-	---	---
TC16000525.hg.1	-0.8291	-5.2259	3.65E-05	7.61E-03	6.1977	6.3862	5.0128	chr16	66995132	67009052	+	NM_001185176	CES3
TC04001500.hg.1	-3.4233	-5.2175	3.73E-05	7.67E-03	7.7806	9.5233	2.8036	chr4	120238405	120243545	-	NM_000134	FABP2
TC11002494.hg.1	-2.6855	-5.2165	3.74E-05	7.67E-03	8.3252	7.2058	4.6150	chr11	1096364	1097364	+	---	---
TC05000563.hg.1	-0.6195	-5.1840	4.03E-05	8.00E-03	6.8616	6.8888	6.9354	chr5	118373467	118584833	+	NM_005509	DMXL1
TC06003507.hg.1	-1.0646	-5.1723	4.14E-05	8.08E-03	4.6590	3.8771	3.8943	chr6	19802395	19804981	-	---	---
TC10002844.hg.1	-0.9398	-5.1737	4.13E-05	8.08E-03	4.8172	4.8039	5.1429	chr10	118587736	118609222	-	---	---
TC19001261.hg.1	-0.9606	-5.1443	4.42E-05	8.48E-03	7.1867	8.4599	5.8052	chr19	15988834	16008884	-	NM_001082	CYP4F2
TC11000811.hg.1	-0.9353	-5.1153	4.73E-05	8.87E-03	5.2651	6.8059	4.1689	chr11	75428864	75444003	+	BC103877	MOGAT2
TC02002858.hg.1	-0.9908	-5.1045	4.85E-05	8.97E-03	6.1596	5.4424	5.0610	chr2	231972944	231989832	-	NM_000867	HTR2B
TC15000409.hg.1	-0.7208	-5.0748	5.20E-05	9.34E-03	3.6359	5.2556	6.5888	chr15	51973550	52013228	+	NM_013243	SCG3
TC04000806.hg.1	-1.0362	-5.0733	5.22E-05	9.34E-03	7.4766	7.7458	7.3705	chr4	159593277	159630775	+	NM_004453	ETFDH

Appendix 5: Top differentially expressed transcript clusters identified by genome-wide expression profiling

Tx cluster ID	logfc	tstat	pval	adj_pval	$Expr_{base}$	$Expr_{duo}$	$Expr_{stom}$	chr	start	end	str	Tx ID	Tx gene
TC08002619.hg.1	-0.6118	-5.0619	5.36E-05	9.43E-03	7.5931	7.4534	7.2223	chr8	124260690	124287781	-	NM_001017926	ZHX1
TC05003086.hg.1	-0.7437	-5.0586	5.40E-05	9.45E-03	5.7928	6.5715	6.5307	chr5	74343544	74348668	-	---	---
TC21000177.hg.1	-0.9679	-5.0475	5.54E-05	9.60E-03	5.6970	5.5768	4.4822	chr21	40928369	41098012	+	NM_006057	B3GALT5
TC05000684.hg.1	-0.8459	-5.0434	5.60E-05	9.66E-03	5.8032	6.1334	4.8542	chr5	135549736	135557847	+	ENST00000514459	TRPC7-AS1
TC15001473.hg.1	-0.5868	-5.0414	5.62E-05	9.66E-03	6.6968	6.4598	6.4836	chr15	56297911	56299359	-	ENST00000393358	CNOT6LP1
TC04001408.hg.1	-2.6324	-5.0369	5.68E-05	9.74E-03	7.5542	8.3771	3.8637	chr4	100123795	100140694	-	NM_000672	ADH6
TC10001555.hg.1	-1.3092	-5.0211	5.90E-05	9.98E-03	8.9081	8.9212	8.3005	chr10	97951455	98031333	-	NM_001114094	BLNK
TC07002319.hg.1	-0.7644	-5.0189	5.93E-05	9.98E-03	2.7389	2.3025	2.7664	chr7	54624663	54639419	+	---	---

Appendix 6: Illumina TruSeq Adapter Sequences

Sequences of Illumina oligonucleotides (proprietary to Illumina, see copyright notice below) used for conjugation with amplified template for targeted sequencing. Index sequences are 6 bases as underlined. Index numbers 17, 24 and 26 are reserved and are not available for sample indexing.

TruSeq Adapter, Index 1
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>ATCACG</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 2
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>CGATGT</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 3
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>TTAGGC</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 4
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>TGACCA</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 5
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>ACAGTG</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 6
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>GCCAAT</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 7
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>CAGATC</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 8
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>ACTTGA</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 9
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>GATCAG</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 10
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>TAGCTT</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 11
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>GGCTAC</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 12
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>CTTGTA</u>ATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 13
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>AGTCAA</u>CAATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 14
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>AGTTCC</u>GTATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 15
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>ATGTCA</u>GAATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 16
5' GATCGGAAGAGCACACGTCTGAACTCCAGTCAC<u>CCGTCC</u>CGATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 18
5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTCCGCACATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 19
5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTGAAACGATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 20
5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTGGCCTTATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 21
5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTTTCGGAATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 22
5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGTACGTAATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 23
5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCACGAGTGGATATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 25
5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTGATATATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 27
5’ GATCGGAAGAGCACACGTCTGAACTCCAGTCACATTCCTTTATCTCGTATGCCGTCTTCTGCTTG

Appendix 7: Hypermethylation biomarker probes for BE and EAC

Hypermethylated probes for identification of intervention requiring disease with respect to normal squamous epithelium

799 probes (top 200, greatest Δβ shown), hypermethylated in HGD and EAC in comparison to N were identified using stringent cut-off criteria (minimal methylation in control tissues and normal peripheral blood). Gencode v19 was used for annotation[204], gene name protein-coding when not specified. logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, β_{base}: average methylation in baseline samples, here, normal squamous epithelium, Δβ: difference in average methylation between comparison groups, here, between HGD-EAC and N, β_{duo}: average methylation in duodenal epithelium, β_{stom}: average methylation in proximal stomach epithelium, β_{blood}: average methylation in peripheral blood from disease-free patients. Listed in order of decreasing Δβ. Entries in grey correspond to target region genes validated by targeted amplicon sequencing (from N v Any disease (intersection between N v BE and N v HGD-EAC) as well as N v HGD-EAC comparison groups). Note that region selection for validation included criteria pertaining to identification of aberrant methylation across groups of probes (within 300bp, at least 2 differentially methylated probes required) as well as region suitability for BSP assay design. Hence some of the top aberrantly hypermethylated individual probes listed here were not part of regions selected for validation.

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
AUTS2	chr7	69064092	69064094	cg17027195	8.5872	20.4221	3.06E-15	1.77E-12	0.0110	0.7995	0.0215	0.0101	0.0110
SCOC	chr4	141295028	141295030	cg21986225	8.2781	15.8192	4.55E-13	6.44E-11	0.0104	0.7550	0.0275	0.0130	0.0100
MTRR	chr5	7850202	7850204	cg12539796	8.1530	17.7910	4.64E-14	1.22E-11	0.0107	0.7445	0.0427	0.0260	0.0134
DNAJC6	chr1	65731431	65731433	cg26615127	6.1030	15.3062	8.58E-13	1.03E-10	0.0503	0.7343	0.0546	0.0445	0.0434
HMX3	chr10	124895447	124895449	cg18685408	6.9471	12.4102	4.45E-11	2.25E-09	0.0241	0.7290	0.0402	0.0119	0.0108
PRDM2	chr1	14026583	14026585	cg00922376	8.0579	24.6511	7.09E-17	1.28E-13	0.0105	0.7289	0.0157	0.0162	0.0100
ZNF790	chr19	37329089	37329091	cg10734240	7.1385	14.4022	2.74E-12	2.50E-10	0.0201	0.7231	0.0260	0.0496	0.0258
N/A	chr1	76080726	76080728	cg07923233	7.6957	12.0282	7.90E-11	3.57E-09	0.0132	0.7218	0.0697	0.0283	0.0129
RUNDC3B	chr7	87257537	87257539	cg18542829	7.9486	30.1483	1.20E-18	6.94E-15	0.0109	0.7207	0.0203	0.0148	0.0105
SYT9	chr11	7273147	7273149	cg18560328	7.0229	17.2739	8.26E-14	1.86E-11	0.0214	0.7182	0.0781	0.0428	0.0123
IGDCC3	chr15	65670303	65670305	cg01107006	7.4118	13.6773	7.27E-12	5.34E-10	0.0152	0.7090	0.0340	0.0172	0.0100
N/A	chr1	76081961	76081963	cg27547954	7.8052	17.1992	8.99E-14	1.97E-11	0.0114	0.7086	0.0534	0.0100	0.0111
LMO1	chr11	8284745	8284747	cg21842523	5.8796	15.3362	8.26E-13	1.01E-10	0.0507	0.7080	0.0649	0.0613	0.0180

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
C11orf96	chr11	43963906	43963908	cg20062650	7.2485	12.0690	7.42E-11	3.39E-09	0.0169	0.7066	0.0396	0.0433	0.0125
GDF10	chr10	48438723	48438725	cg04110601	7.4536	18.3007	2.67E-14	8.16E-12	0.0145	0.7061	0.0959	0.0322	0.0164
HMX3	chr10	124895473	124895475	cg00560482	6.4198	14.7924	1.65E-12	1.70E-10	0.0316	0.7049	0.0529	0.0374	0.0181
RUNDC3B	chr7	87257672	87257674	cg17960051	7.8504	32.0346	3.49E-19	2.87E-15	0.0107	0.7041	0.0110	0.0229	0.0100
TMEM178B	chr7	140773904	140773906	cg07028821	7.3338	22.2452	5.57E-16	5.22E-13	0.0154	0.7009	0.0884	0.0460	0.0126
PLXNA4	chr7	132262352	132262354	cg07258916	7.2773	17.7944	4.63E-14	1.22E-11	0.0160	0.7002	0.0292	0.0272	0.0158
HS3ST4	chr16	25703527	25703529	cg27014135	7.6340	25.9176	2.58E-17	6.02E-14	0.0123	0.7000	0.0603	0.0100	0.0100
CTD-2245F17.3 lincRNA	chr19	53700526	53700528	cg23021477	7.2922	19.1884	1.05E-14	4.30E-12	0.0158	0.6997	0.0142	0.0530	0.0149
PURG	chr8	30890619	30890621	cg18324126	7.5802	16.6990	1.60E-13	3.00E-11	0.0127	0.6990	0.0399	0.0179	0.0100
T protein_coding	chr6	166582309	166582311	cg06073449	7.1279	13.4022	1.07E-11	7.22E-10	0.0177	0.6981	0.0537	0.0161	0.0100
ZNF662	chr3	42947689	42947691	cg24384244	7.2728	13.1744	1.47E-11	9.31E-10	0.0156	0.6946	0.0631	0.0468	0.0131
ZNF829	chr19	37407215	37407217	cg20680720	7.7261	33.5780	1.33E-19	1.36E-15	0.0111	0.6935	0.0972	0.0133	0.0119
KCNMA1	chr10	79397454	79397456	cg09405661	6.4890	12.2274	5.84E-11	2.80E-09	0.0276	0.6904	0.0388	0.0468	0.0313
PURG	chr8	30890616	30890618	cg01755467	7.7055	15.7107	5.20E-13	7.12E-11	0.0110	0.6887	0.0229	0.0108	0.0100
RUNDC3B	chr7	87257073	87257075	cg04710402	7.1399	21.0156	1.73E-15	1.17E-12	0.0166	0.6876	0.0688	0.0354	0.0116
CLSTN2	chr3	139654263	139654265	cg22610211	7.6646	9.9617	2.28E-09	5.89E-08	0.0113	0.6872	0.0332	0.0116	0.0100
NDRG4	chr16	58497814	58497816	cg02040433	7.7941	28.7004	3.27E-18	1.42E-14	0.0102	0.6856	0.0909	0.0103	0.0105
PODN	chr1	53527664	53527666	cg01394819	7.2886	14.8584	1.51E-12	1.59E-10	0.0147	0.6851	0.0316	0.0205	0.0133
EPHA7	chr6	94129626	94129628	cg06740629	6.0442	7.7775	1.38E-07	1.98E-06	0.0377	0.6834	0.0787	0.0637	0.0273
KIRREL	chr1	157963810	157963812	cg12612104	6.6430	13.6640	7.41E-12	5.43E-10	0.0233	0.6811	0.0407	0.0477	0.0200
KCNQ3	chr8	133493423	133493425	cg22308501	6.9964	9.5663	4.59E-09	1.06E-07	0.0178	0.6808	0.0143	0.0265	0.0136
SDC2	chr8	97507560	97507562	cg10292139	5.8141	12.5914	3.40E-11	1.81E-09	0.0439	0.6771	0.0413	0.0514	0.0471
HMX3	chr10	124895460	124895462	cg14663510	5.8692	13.5466	8.71E-12	6.18E-10	0.0416	0.6756	0.0800	0.0441	0.0275
WIPF1	chr2	175546915	175546917	cg26831241	6.8703	16.9994	1.13E-13	2.31E-11	0.0188	0.6725	0.0150	0.0320	0.0233
CELF2	chr10	11059726	11059728	cg12356890	6.3607	9.3509	6.76E-09	1.48E-07	0.0276	0.6725	0.0313	0.0604	0.0266
NTNG1	chr1	107683714	107683716	cg11396157	6.8935	19.2390	9.97E-15	4.15E-12	0.0183	0.6707	0.0924	0.0149	0.0112
RP4-555D20.4 lincRNA	chr3	44040810	44040812	cg27606567	6.6427	22.5873	4.10E-16	4.22E-13	0.0220	0.6701	0.0981	0.0477	0.0164
BMP3	chr4	81952023	81952025	cg26917673	6.5372	22.0933	6.39E-16	5.68E-13	0.0238	0.6699	0.0141	0.0373	0.0121
N/A	chr2	105479053	105479055	cg11014373	6.3844	15.1067	1.10E-12	1.26E-10	0.0267	0.6697	0.0925	0.0612	0.0100

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
PRDM2	chr1	14026604	14026606	cg09765994	6.5614	18.6329	1.87E-14	6.40E-12	0.0233	0.6696	0.0246	0.0293	0.0144
RET	chr10	43572064	43572066	cg05621401	6.8539	11.6842	1.34E-10	5.51E-09	0.0187	0.6689	0.0343	0.0309	0.0100
ADHFE1	chr8	67344587	67344589	cg18065361	7.5235	19.0847	1.17E-14	4.62E-12	0.0113	0.6664	0.0618	0.0139	0.0100
WHAMMP2 pseudogene	chr15	28983116	28983118	cg11794877	7.3934	29.8372	1.48E-18	8.22E-15	0.0124	0.6663	0.0177	0.0105	0.0100
TNFSF11	chr13	43148934	43148936	cg26224671	6.6724	13.8532	5.72E-12	4.43E-10	0.0208	0.6633	0.0520	0.0379	0.0235
RP11-1055B8.7	chr17	79373380	79373382	cg16560774	7.0204	14.3365	2.99E-12	2.68E-10	0.0161	0.6632	0.0837	0.0404	0.0100
CPLX2	chr5	175223981	175223983	cg07295964	6.8016	13.4048	1.06E-11	7.20E-10	0.0188	0.6621	0.0987	0.0529	0.0145
KCNMA1	chr10	79398054	79398056	cg24113782	6.6829	8.7793	1.95E-08	3.67E-07	0.0205	0.6620	0.0144	0.0185	0.0199
PURG	chr8	30890631	30890633	cg23139473	6.9123	12.6829	2.98E-11	1.63E-09	0.0172	0.6605	0.0216	0.0190	0.0121
CELF2	chr10	11059732	11059734	cg11472279	6.3952	9.2650	7.91E-09	1.69E-07	0.0252	0.6597	0.0335	0.0552	0.0145
BVES	chr6	105584550	105584552	cg14159026	6.3043	19.9856	4.70E-15	2.41E-12	0.0270	0.6596	0.0835	0.0445	0.0184
H2AFY2	chr10	71813352	71813354	cg19726179	6.7892	17.4173	7.03E-14	1.65E-11	0.0187	0.6591	0.0607	0.0322	0.0134
SNCA	chr4	90758796	90758798	cg20776829	7.2147	11.1523	3.11E-10	1.10E-08	0.0136	0.6589	0.0843	0.0824	0.0111
GATA5	chr20	61051340	61051342	cg16714055	5.9028	13.1064	1.62E-11	1.00E-09	0.0367	0.6581	0.0780	0.0354	0.0332
AC079776.1 lincRNA	chr2	130635165	130635167	cg06480249	7.6369	17.1478	9.53E-14	2.05E-11	0.0100	0.6578	0.0100	0.0114	0.0103
T protein_coding	chr6	166582205	166582207	cg19675288	5.7797	15.9228	4.01E-13	5.85E-11	0.0403	0.6574	0.0837	0.0580	0.0219
CELF2	chr10	11060651	11060653	cg13950829	5.4920	9.9944	2.16E-09	5.61E-08	0.0511	0.6568	0.0686	0.0940	0.0259
RP11-139I14.2 lincRNA	chr1	156358230	156358232	cg12062819	7.4740	14.0662	4.29E-12	3.55E-10	0.0111	0.6551	0.0813	0.0116	0.0103
KIF5C	chr2	149632704	149632706	cg14986699	7.5956	10.8637	4.97E-10	1.63E-08	0.0102	0.6551	0.0397	0.0243	0.0123
NDRG4	chr16	58497800	58497802	cg08791131	7.5531	24.6331	7.19E-17	1.29E-13	0.0105	0.6551	0.0446	0.0100	0.0100
EFCC1/KIAA1257	chr3	128720432	128720434	cg20506715	5.4470	12.9964	1.89E-11	1.13E-09	0.0524	0.6546	0.0912	0.0655	0.0314
GATA5	chr20	61051431	61051433	cg08568720	6.3908	11.6262	1.47E-10	5.93E-09	0.0246	0.6545	0.0786	0.0195	0.0155
ENOX1	chr13	44361219	44361221	cg01118451	7.0402	12.3084	5.18E-11	2.53E-09	0.0151	0.6537	0.0283	0.0272	0.0102
SCOC	chr4	141294680	141294682	cg20366601	7.4083	10.6187	7.45E-10	2.28E-08	0.0115	0.6534	0.0135	0.0106	0.0114
NA	chr19	40314926	40314928	cg06936564	5.0145	7.5515	2.18E-07	2.95E-06	0.0774	0.6532	0.0971	0.0854	0.0334
PCBP3	chr21	47063284	47063286	cg01145054	7.2598	7.5567	2.16E-07	2.93E-06	0.0128	0.6522	0.0190	0.0144	0.0100
ANKLE1	chr19	17392769	17392771	cg27101125	5.9897	14.9595	1.33E-12	1.45E-10	0.0329	0.6509	0.0711	0.0451	0.0207
SHD	chr19	4279440	4279442	cg26646370	4.9837	13.7830	6.29E-12	4.77E-10	0.0777	0.6495	0.0629	0.0835	0.0692

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
TRBJ2-1	chr7	142494913	142494915	cg21621906	7.3948	22.0045	6.92E-16	6.09E-13	0.0114	0.6479	0.0846	0.0111	0.0104
KIAA0226L	chr13	46961582	46961584	cg23833588	6.4562	13.2700	1.28E-11	8.38E-10	0.0225	0.6462	0.0126	0.0281	0.0100
DRD4	chr11	637174	637176	cg12928379	5.5851	18.8822	1.44E-14	5.30E-12	0.0444	0.6459	0.0492	0.0361	0.0221
ZNF461	chr19	37157994	37157996	cg27545919	5.8829	12.7045	2.88E-11	1.59E-09	0.0349	0.6459	0.0625	0.0437	0.0204
OBSL1	chr2	220417248	220417250	cg11531021	6.4123	11.9363	9.08E-11	4.01E-09	0.0232	0.6458	0.0239	0.0271	0.0133
KCNQ3	chr8	133493170	133493172	cg13070215	7.4738	8.7276	2.15E-08	3.99E-07	0.0106	0.6457	0.0110	0.0142	0.0115
ZNF790	chr19	37329380	37329382	cg18876786	7.3732	20.8873	1.96E-15	1.28E-12	0.0114	0.6452	0.0185	0.0190	0.0116
TRPC4	chr13	38443949	38443951	cg19275632	6.1212	10.0497	1.96E-09	5.17E-08	0.0288	0.6447	0.0710	0.0368	0.0286
N/A	chr8	25909325	25909327	cg00829961	6.3733	10.0520	1.95E-09	5.15E-08	0.0237	0.6445	0.0914	0.0175	0.0177
SAMD3/TMEM200A	chr6	130686696	130686698	cg20334627	6.6517	9.9826	2.20E-09	5.71E-08	0.0192	0.6440	0.0248	0.0528	0.0121
KCNQ3	chr8	133492475	133492477	cg27016990	7.4037	40.2134	3.31E-21	1.90E-16	0.0111	0.6435	0.0639	0.0271	0.0113
GALNT14	chr2	31360692	31360694	cg21583226	6.3703	19.6874	6.33E-15	2.99E-12	0.0236	0.6432	0.0544	0.0469	0.0135
TRBJ2-1	chr7	142494952	142494954	cg08430489	6.1580	17.9820	3.77E-14	1.05E-11	0.0276	0.6422	0.1000	0.0409	0.0217
GATA5	chr20	61051347	61051349	cg11982072	5.5588	13.4094	1.05E-11	7.17E-10	0.0440	0.6406	0.0789	0.0315	0.0347
DTX3	chr12	57998785	57998787	cg17730484	5.3911	10.8902	4.76E-10	1.57E-08	0.0503	0.6395	0.0488	0.0563	0.0465
CNTFR	chr9	34589710	34589712	cg13518298	5.8620	18.4328	2.32E-14	7.38E-12	0.0342	0.6391	0.0343	0.0417	0.0117
EPB41L3	chr18	5629682	5629684	cg03718824	5.9428	15.8703	4.28E-13	6.14E-11	0.0320	0.6384	0.0433	0.0883	0.0371
ENOX1	chr13	44361211	44361213	cg24098938	7.0579	12.7327	2.77E-11	1.54E-09	0.0139	0.6383	0.0191	0.0117	0.0127
EHD3	chr2	31456774	31456776	cg18444347	6.2639	15.2463	9.25E-13	1.10E-10	0.0250	0.6381	0.0623	0.0306	0.0207
TMEM178B	chr7	140773735	140773737	cg16633901	5.7002	19.6729	6.42E-15	3.02E-12	0.0387	0.6380	0.0804	0.0766	0.0398
MED12L	chr3	150803668	150803670	cg21401219	5.4639	12.7528	2.69E-11	1.50E-09	0.0466	0.6365	0.0643	0.0711	0.0238
EOMES	chr3	27763101	27763103	cg24434959	6.1404	20.5374	2.74E-15	1.64E-12	0.0272	0.6365	0.0278	0.0403	0.0228
CAMK2B	chr7	44364924	44364926	cg17035091	7.1596	28.6349	3.42E-18	1.47E-14	0.0127	0.6354	0.0132	0.0193	0.0128
SGCE	chr7	94284431	94284433	cg19734015	5.6063	7.4622	2.62E-07	3.46E-06	0.0411	0.6351	0.0857	0.0335	0.0133
GRID2	chr4	93226379	93226381	cg02468050	6.1239	13.6981	7.07E-12	5.22E-10	0.0274	0.6351	0.0751	0.0386	0.0211
TLL1	chr4	166794785	166794787	cg24521633	6.5363	15.1003	1.11E-12	1.27E-10	0.0200	0.6349	0.0420	0.0286	0.0146
RUNDC3B	chr7	87257766	87257768	cg15414833	5.5316	18.2767	2.74E-14	8.34E-12	0.0435	0.6344	0.0433	0.0521	0.0376
PRDM2	chr1	14026589	14026591	cg05346841	7.1979	22.6269	3.96E-16	4.10E-13	0.0123	0.6341	0.0208	0.0277	0.0100
OSTM1/NR2E1	chr6	108487159	108487161	cg18918349	5.9064	9.2492	8.14E-09	1.74E-07	0.0322	0.6339	0.0850	0.0433	0.0290
N/A	chr7	1709949	1709951	cg10669265	6.3108	16.2049	2.86E-13	4.58E-11	0.0236	0.6335	0.0870	0.0384	0.0144
MPPED2	chr11	30607067	30607069	cg11855526	6.2535	25.9300	2.56E-17	6.00E-14	0.0245	0.6324	0.0494	0.0373	0.0153

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
FOXI3	chr2	88752321	88752323	cg23511613	7.4235	32.2323	3.08E-19	2.57E-15	0.0104	0.6322	0.0989	0.0125	0.0105
PRKCQ	chr10	6622554	6622556	cg19910780	6.3292	15.7237	5.11E-13	7.02E-11	0.0230	0.6314	0.0259	0.0255	0.0227
STAC2	chr17	37382172	37382174	cg07280482	6.2057	14.7219	1.80E-12	1.82E-10	0.0252	0.6309	0.0982	0.0424	0.0187
PRR5	chr22	45064470	45064472	cg04480903	6.5405	13.1934	1.43E-11	9.13E-10	0.0196	0.6308	0.0697	0.0224	0.0161
L3MBTL4	chr18	6414973	6414975	cg18556788	6.3332	13.5215	9.02E-12	6.35E-10	0.0229	0.6307	0.0513	0.0426	0.0172
N/A	chr7	1704583	1704585	cg06700935	6.2979	15.9453	3.90E-13	5.74E-11	0.0233	0.6295	0.0421	0.0310	0.0106
GAMT	chr19	1401309	1401311	cg24102241	7.2337	17.6728	5.29E-14	1.34E-11	0.0117	0.6294	0.0432	0.0169	0.0104
EVL	chr14	100438736	100438738	cg21164440	6.0070	17.7565	4.82E-14	1.26E-11	0.0291	0.6294	0.0312	0.0287	0.0197
ZNF665	chr19	53696648	53696650	cg01620580	6.7938	15.9956	3.67E-13	5.50E-11	0.0161	0.6291	0.0731	0.0154	0.0161
SIX2	chr2	45237688	45237690	cg02711647	5.5078	9.7187	3.50E-09	8.45E-08	0.0431	0.6288	0.0893	0.0186	0.0228
SH3RF3	chr2	109745827	109745829	cg09392940	6.7878	18.8365	1.51E-14	5.45E-12	0.0162	0.6285	0.0167	0.0215	0.0142
AUTS2	chr7	69063474	69063476	cg21591173	6.0799	10.2520	1.38E-09	3.86E-08	0.0274	0.6282	0.0216	0.0285	0.0229
PRDM2	chr1	14026617	14026619	cg07596209	7.0365	19.9890	4.68E-15	2.41E-12	0.0135	0.6282	0.0171	0.0151	0.0106
FOXL2	chr3	138666117	138666119	cg06186698	6.3491	14.3611	2.89E-12	2.61E-10	0.0222	0.6274	0.0860	0.0347	0.0218
NRG1	chr8	32406927	32406929	cg06926782	6.4958	13.7397	6.68E-12	4.99E-10	0.0199	0.6274	0.0891	0.0341	0.0164
MED12L	chr3	150804062	150804064	cg24255728	5.6107	19.4389	8.13E-15	3.56E-12	0.0393	0.6274	0.0256	0.0466	0.0223
TBX21	chr17	45810864	45810866	cg09775582	7.2370	30.7668	7.95E-19	5.30E-15	0.0116	0.6266	0.0874	0.0100	0.0100
ADCY5	chr3	123167769	123167771	cg02978184	6.3752	10.9513	4.31E-10	1.45E-08	0.0217	0.6262	0.0367	0.0305	0.0125
ELOVL5	chr6	53213870	53213872	cg21195414	7.2580	20.5902	2.60E-15	1.57E-12	0.0114	0.6261	0.0103	0.0820	0.0101
N/A	chr7	1705348	1705350	cg06134410	5.9519	19.0129	1.26E-14	4.86E-12	0.0298	0.6256	0.0489	0.0346	0.0328
ST8SIA1	chr12	22488228	22488230	cg24239882	5.2033	8.6129	2.67E-08	4.81E-07	0.0543	0.6246	0.0999	0.0641	0.0769
CALY	chr10	135139099	135139101	cg26583481	6.7123	15.6892	5.33E-13	7.25E-11	0.0167	0.6236	0.0408	0.0359	0.0180
ZNF790/ZNF345	chr19	37341733	37341735	cg12033943	5.5624	16.3561	2.39E-13	4.01E-11	0.0401	0.6235	0.0227	0.0355	0.0389
POU3F2	chr6	99283200	99283202	cg16202970	6.4171	15.0950	1.12E-12	1.27E-10	0.0207	0.6227	0.0662	0.0159	0.0106
FAM78B	chr1	166134281	166134283	cg21868774	6.9192	13.0140	1.84E-11	1.11E-09	0.0143	0.6226	0.0172	0.0262	0.0128
DTX3	chr12	57998761	57998763	cg11654179	5.2805	12.0104	8.11E-11	3.65E-09	0.0500	0.6216	0.0527	0.0691	0.0203
NEUROG1	chr5	134871965	134871967	cg07035503	6.8408	16.8405	1.36E-13	2.65E-11	0.0150	0.6215	0.0892	0.0438	0.0102
KIF6	chr6	39692937	39692939	cg26867890	5.5147	13.5583	8.57E-12	6.10E-10	0.0411	0.6212	0.0929	0.0490	0.0138
PROKR1	chr2	68870811	68870813	cg13451280	7.2977	9.6473	3.97E-09	9.42E-08	0.0108	0.6208	0.0316	0.0130	0.0101
ELOVL5	chr6	53213158	53213160	cg13529912	7.2005	9.3349	6.96E-09	1.52E-07	0.0115	0.6202	0.0100	0.0121	0.0100
ARMC4	chr10	28288002	28288004	cg13397820	5.7467	11.3542	2.25E-10	8.42E-09	0.0339	0.6196	0.0778	0.0594	0.0122

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
GRIN3A	chr9	104501029	104501031	cg18794577	6.0871	15.9796	3.75E-13	5.58E-11	0.0261	0.6195	0.0599	0.0447	0.0320
PRDM2	chr1	14026816	14026818	cg09826587	7.2065	20.6811	2.38E-15	1.48E-12	0.0114	0.6191	0.0116	0.0182	0.0100
ADARB2	chr10	1779834	1779836	cg02899206	5.6091	17.8687	4.26E-14	1.15E-11	0.0377	0.6191	0.0275	0.0456	0.0377
SNX32	chr11	65601331	65601333	cg11198128	7.3460	32.3319	2.89E-19	2.52E-15	0.0103	0.6191	0.0677	0.0100	0.0100
ELOVL5	chr6	53213027	53213029	cg27599958	6.7477	11.2512	2.65E-10	9.67E-09	0.0159	0.6190	0.0142	0.0301	0.0151
BEND5	chr1	49242518	49242520	cg16573178	6.3514	26.3634	1.83E-17	4.72E-14	0.0213	0.6188	0.0199	0.0295	0.0128
CTD-2319I12.5 lincRNA	chr17	58217356	58217358	cg14689219	5.2668	8.4718	3.50E-08	6.07E-07	0.0497	0.6185	0.0591	0.0328	0.0100
HMX3	chr10	124895637	124895639	cg11628754	6.0534	16.4579	2.12E-13	3.69E-11	0.0266	0.6184	0.0735	0.0308	0.0165
FGF14-AS2 lincRNA	chr13	103047052	103047054	cg16398329	7.2364	35.2351	4.98E-20	6.74E-16	0.0111	0.6182	0.0942	0.0141	0.0100
N/A	chr10	131769580	131769582	cg15344419	5.2108	20.9298	1.88E-15	1.25E-12	0.0519	0.6178	0.0818	0.0333	0.0230
CLSTN2	chr3	139654244	139654246	cg08853659	6.2904	18.3889	2.43E-14	7.62E-12	0.0222	0.6174	0.0623	0.0257	0.0169
OBSL1	chr2	220417029	220417031	cg02527669	5.5641	16.4478	2.14E-13	3.73E-11	0.0388	0.6173	0.0794	0.0595	0.0323
SAMD3/TMEM200A	chr6	130686707	130686709	cg26483578	6.4653	8.9200	1.50E-08	2.94E-07	0.0194	0.6171	0.0249	0.0439	0.0180
MAD2L2	chr1	11752161	11752163	cg02397176	6.8389	12.4094	4.45E-11	2.25E-09	0.0148	0.6170	0.0896	0.0257	0.0164
N/A	chr5	1667642	1667644	cg15658945	5.0091	12.0373	7.79E-11	3.53E-09	0.0615	0.6169	0.0476	0.0507	0.0408
VENTX	chr10	135050778	135050780	cg18002909	5.4405	12.4759	4.03E-11	2.08E-09	0.0424	0.6154	0.0991	0.0752	0.0427
EIF4E3	chr3	71803557	71803559	cg10172415	6.5836	17.4694	6.63E-14	1.58E-11	0.0177	0.6152	0.0481	0.0264	0.0118
EPDR1	chr7	37960901	37960903	cg08608193	5.6183	10.1348	1.69E-09	4.56E-08	0.0366	0.6147	0.0186	0.0399	0.0180
CNTFR	chr9	34589386	34589388	cg13578447	6.4554	16.6664	1.66E-13	3.09E-11	0.0194	0.6147	0.0303	0.0299	0.0261
N/A	chr1	111098149	111098151	cg02585459	5.0450	11.5918	1.55E-10	6.19E-09	0.0587	0.6145	0.0499	0.0747	0.0470
KIF6	chr6	39693359	39693361	cg01138981	6.9366	12.5875	3.42E-11	1.82E-09	0.0136	0.6144	0.0204	0.0152	0.0103
SH3RF3	chr2	109745811	109745813	cg12518410	7.1217	18.8779	1.45E-14	5.30E-12	0.0119	0.6142	0.0100	0.0146	0.0113
CLDN11	chr3	170136498	170136500	cg21389743	6.2786	14.0673	4.28E-12	3.55E-10	0.0220	0.6139	0.0270	0.0352	0.0174
H2AFY2	chr10	71812611	71812613	cg26166804	5.5202	17.9283	4.00E-14	1.10E-11	0.0394	0.6135	0.0627	0.0514	0.0466
TUB	chr11	8103016	8103018	cg09498146	6.3395	15.5852	6.06E-13	7.98E-11	0.0210	0.6132	0.0919	0.0227	0.0162
B3GAT2	chr6	71666681	71666683	cg16556906	6.7127	23.1880	2.43E-16	2.87E-13	0.0158	0.6121	0.0858	0.0185	0.0156
H2AFY2	chr10	71812595	71812597	cg17163751	5.4222	18.2084	2.95E-14	8.82E-12	0.0422	0.6117	0.0622	0.0565	0.0502
MED12L	chr3	150804057	150804059	cg12217936	7.0489	27.6807	6.81E-18	2.40E-14	0.0124	0.6113	0.0100	0.0100	0.0100
BMP3	chr4	81951955	81951957	cg22403273	6.0836	13.4772	9.59E-12	6.65E-10	0.0251	0.6111	0.0170	0.0442	0.0184
SLIT2	chr4	20253513	20253515	cg13755796	5.5292	11.7199	1.27E-10	5.27E-09	0.0386	0.6109	0.0376	0.0390	0.0126

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
ALX4	chr11	44330902	44330904	cg26365854	6.7108	20.8148	2.10E-15	1.34E-12	0.0158	0.6107	0.0915	0.0128	0.0100
NEUROG1	chr5	134871806	134871808	cg22630755	6.9868	18.2025	2.97E-14	8.84E-12	0.0129	0.6105	0.0726	0.0205	0.0111
PRKCB	chr16	23847239	23847241	cg00735962	6.0049	17.8395	4.40E-14	1.17E-11	0.0266	0.6105	0.0183	0.0194	0.0144
GNG4	chr1	235813443	235813445	cg20285514	6.5341	10.7241	6.25E-10	1.98E-08	0.0179	0.6103	0.0193	0.0318	0.0177
CD1D	chr1	158150648	158150650	cg12124922	5.7274	17.0940	1.01E-13	2.14E-11	0.0328	0.6099	0.0273	0.0600	0.0211
ATP6V1B1	chr2	71192261	71192263	cg10598816	4.8767	10.0052	2.12E-09	5.52E-08	0.0663	0.6095	0.0538	0.0642	0.0510
ZNF790/ZNF345	chr19	37341869	37341871	cg03100040	5.1806	12.4490	4.20E-11	2.14E-09	0.0507	0.6087	0.0394	0.0585	0.0357
KCNQ3	chr8	133493433	133493435	cg04396550	5.3879	10.1843	1.55E-09	4.24E-08	0.0427	0.6087	0.0191	0.0219	0.0139
CELF2	chr10	11059576	11059578	cg23858040	7.2090	12.0365	7.80E-11	3.54E-09	0.0109	0.6080	0.0245	0.0257	0.0100
FGFR1	chr8	38325949	38325951	cg27063138	5.5583	9.5182	5.00E-09	1.14E-07	0.0371	0.6080	0.0231	0.0304	0.0254
KIAA1211L	chr2	99553441	99553443	cg04787024	5.6110	14.8013	1.63E-12	1.68E-10	0.0356	0.6077	0.0948	0.0607	0.0328
ZNF569/ZNF570	chr19	37957994	37957996	cg14142713	5.2517	23.4589	1.92E-16	2.47E-13	0.0475	0.6076	0.0559	0.0363	0.0506
EPDR1	chr7	37960872	37960874	cg10876076	4.8110	10.3249	1.22E-09	3.47E-08	0.0694	0.6074	0.0519	0.0749	0.0246
PPP1R14A	chr19	38747200	38747202	cg13564825	5.9802	12.6420	3.16E-11	1.71E-09	0.0267	0.6072	0.0655	0.0423	0.0197
CWH43	chr4	48988252	48988254	cg22930650	5.5056	12.3030	5.22E-11	2.55E-09	0.0384	0.6064	0.0590	0.0489	0.0226
MME	chr3	154797916	154797918	cg23209255	6.2001	9.6999	3.62E-09	8.69E-08	0.0225	0.6064	0.0199	0.0175	0.0106
RP11-1055B8.7	chr17	79373672	79373674	cg07219667	5.6287	11.1415	3.16E-10	1.12E-08	0.0349	0.6064	0.0657	0.0377	0.0269
CD38	chr4	15780305	15780307	cg15994026	6.1665	10.3768	1.12E-09	3.22E-08	0.0228	0.6034	0.0438	0.0444	0.0103
EPHA6	chr3	96533289	96533291	cg15093079	6.7418	21.4263	1.18E-15	9.09E-13	0.0149	0.6034	0.0918	0.0439	0.0157
ZNF71	chr19	57106537	57106539	cg22995684	6.0414	21.7878	8.43E-16	7.16E-13	0.0250	0.6030	0.0632	0.0298	0.0258
ISM1	chr20	13201455	13201457	cg18993918	6.3228	15.0757	1.15E-12	1.29E-10	0.0202	0.6022	0.0862	0.0655	0.0149
NTNG1	chr1	107683186	107683188	cg07005523	5.9594	11.9949	8.31E-11	3.72E-09	0.0265	0.6022	0.0731	0.0364	0.0297
WIPF1	chr2	175547398	175547400	cg25075147	6.4759	22.0444	6.68E-16	5.92E-13	0.0180	0.6019	0.0148	0.0271	0.0100
GATA5	chr20	61051422	61051424	cg24500900	6.6628	13.4728	9.65E-12	6.68E-10	0.0157	0.6014	0.0336	0.0216	0.0114
TSLP	chr5	110408996	110408998	cg24994173	6.9507	19.5733	7.10E-15	3.26E-12	0.0127	0.6013	0.0515	0.0312	0.0115
PRKG1	chr10	52833908	52833910	cg04884011	5.7674	11.0806	3.49E-10	1.21E-08	0.0305	0.6007	0.0803	0.0332	0.0275
FADS2/FADS1	chr11	61595484	61595486	cg23760165	7.0829	30.4147	1.00E-18	6.36E-15	0.0115	0.6007	0.0107	0.0141	0.0107
KCTD8	chr4	44450880	44450882	cg11258164	5.8484	11.9653	8.69E-11	3.86E-09	0.0285	0.6001	0.0438	0.0260	0.0315
CNTNAP2	chr7	145813007	145813009	cg09571420	6.2109	19.4319	8.19E-15	3.58E-12	0.0217	0.6001	0.0950	0.0307	0.0107
PTHLH	chr12	28122370	28122372	cg19083459	6.8207	11.1136	3.31E-10	1.16E-08	0.0139	0.6000	0.0475	0.0156	0.0122
L3MBTL4	chr18	6414957	6414959	cg22058122	6.1714	19.7903	5.71E-15	2.76E-12	0.0223	0.5999	0.0411	0.0381	0.0213

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
ADCY5	chr3	123167506	123167508	cg15993383	5.7653	12.3511	4.86E-11	2.41E-09	0.0304	0.5999	0.0412	0.0540	0.0210
DCHS1	chr11	6677086	6677088	cg21012362	6.5492	11.4311	1.99E-10	7.61E-09	0.0169	0.5998	0.0133	0.0227	0.0192
GSC	chr14	95236122	95236124	cg10042799	6.8953	17.6975	5.15E-14	1.32E-11	0.0131	0.5997	0.0435	0.0279	0.0100
ZNF569/ZNF570	chr19	37957996	37957998	cg03884783	5.0340	23.7619	1.48E-16	2.08E-13	0.0546	0.5996	0.0610	0.0474	0.0596
NAV2	chr11	19735700	19735702	cg20686479	6.2210	10.7094	6.41E-10	2.01E-08	0.0215	0.5994	0.0751	0.0331	0.0126
CXCL12	chr10	44880561	44880563	cg26267854	6.0079	14.7073	1.84E-12	1.84E-10	0.0252	0.5990	0.0977	0.0440	0.0141
MPPED1	chr11	30607359	30607361	cg24185576	4.6608	20.9207	1.89E-15	1.25E-12	0.0756	0.5985	0.0853	0.0656	0.0592

Hypermethylated probes for identification of Barrett's esophagus mucosa with respect to normal squamous epithelium

235 probes (all shown), hypermethylated in non-dysplastic BE in comparison to N were identified using stringent cut-off criteria (minimal methylation in control tissues and normal peripheral blood). Gencode v19 used for annotation[204], gene name protein-coding when not specified. logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, β_{base}: average methylation in baseline samples, here, normal squamous epithelium, Δβ: difference in average methylation between comparison groups, here, between non-dysplastic BE and N, β_{duo}: average methylation in duodenal epithelium, β_{stom}: average methylation in proximal stomach epithelium, β_{blood}: average methylation in peripheral blood from disease-free patients. Listed in order of decreasing Δβ. Entries in grey correspond to target region genes validated by targeted amplicon sequencing (from N v Any disease (intersection between N v BE and N v HGD-EAC) as well as N v HGD-EAC comparison groups). Note that region selection for validation included criteria pertaining to identification of aberrant methylation across groups of probes (within 300bp, at least 2 differentially methylated probes required) as well as region suitability for BSP assay design. Hence some of the top aberrantly hypermethylated individual probes listed here were not part of regions selected for validation.

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
AUTS2	chr7	69064092	69064094	cg17027195	8.5872	20.4221	3.06E-15	1.77E-12	0.0110	0.7995	0.0215	0.0101	0.0110
SCOC	chr4	141295028	141295030	cg21986225	8.2781	15.8192	4.55E-13	6.44E-11	0.0104	0.7550	0.0275	0.0130	0.0100
MTRR	chr5	7850202	7850204	cg12539796	8.1530	17.7910	4.64E-14	1.22E-11	0.0107	0.7445	0.0427	0.0260	0.0134
DNAJC6	chr1	65731431	65731433	cg26615127	6.1030	15.3062	8.58E-13	1.03E-10	0.0503	0.7343	0.0546	0.0445	0.0434
PRDM2	chr1	14026583	14026585	cg00922376	8.0579	24.6511	7.09E-17	1.28E-13	0.0105	0.7289	0.0157	0.0162	0.0100
ZNF790	chr19	37329089	37329091	cg10734240	7.1385	14.4022	2.74E-12	2.50E-10	0.0201	0.7231	0.0260	0.0496	0.0258

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
RUNDC3B	chr7	87257537	87257539	cg18542829	7.9486	30.1483	1.20E-18	6.94E-15	0.0109	0.7207	0.0203	0.0148	0.0105
SYT9	chr11	7273147	7273149	cg18560328	7.0229	17.2739	8.26E-14	1.86E-11	0.0214	0.7182	0.0781	0.0428	0.0123
GDF10	chr10	48438723	48438725	cg04110601	7.4536	18.3007	2.67E-14	8.16E-12	0.0145	0.7061	0.0959	0.0322	0.0164
RUNDC3B	chr7	87257672	87257674	cg17960051	7.8504	32.0346	3.49E-19	2.87E-15	0.0107	0.7041	0.0110	0.0229	0.0100
TMEM178B	chr7	140773904	140773906	cg07028821	7.3338	22.2452	5.57E-16	5.22E-13	0.0154	0.7009	0.0884	0.0460	0.0126
PLXNA4	chr7	132262352	132262354	cg07258916	7.2773	17.7944	4.63E-14	1.22E-11	0.0160	0.7002	0.0292	0.0272	0.0158
HS3ST4	chr16	25703527	25703529	cg27014135	7.6340	25.9176	2.58E-17	6.02E-14	0.0123	0.7000	0.0603	0.0100	0.0100
CTD-2245F17.3 lincRNA	chr19	53700526	53700528	cg23021477	7.2922	19.1884	1.05E-14	4.30E-12	0.0158	0.6997	0.0142	0.0530	0.0149
T protein_coding	chr6	166582309	166582311	cg06073449	7.1279	13.4022	1.07E-11	7.22E-10	0.0177	0.6981	0.0537	0.0161	0.0100
ZNF568	chr19	37407215	37407217	cg20680720	7.7261	33.5780	1.33E-19	1.36E-15	0.0111	0.6935	0.0972	0.0133	0.0119
RUNDC3B	chr7	87257073	87257075	cg04710402	7.1399	21.0156	1.73E-15	1.17E-12	0.0166	0.6876	0.0688	0.0354	0.0116
NDRG4	chr16	58497814	58497816	cg02040433	7.7941	28.7004	3.27E-18	1.42E-14	0.0102	0.6856	0.0909	0.0103	0.0105
KIRREL	chr1	157963810	157963812	cg12612104	6.6430	13.6640	7.41E-12	5.43E-10	0.0233	0.6811	0.0407	0.0477	0.0200
RP4-555D20.4 lincRNA	chr3	44040810	44040812	cg27606567	6.6427	22.5873	4.10E-16	4.22E-13	0.0220	0.6701	0.0981	0.0477	0.0164
BMP3	chr4	81952023	81952025	cg26917673	6.5372	22.0933	6.39E-16	5.68E-13	0.0238	0.6699	0.0141	0.0373	0.0121
N/A	chr2	105479053	105479055	cg11014373	6.3844	15.1067	1.10E-12	1.26E-10	0.0267	0.6697	0.0925	0.0612	0.0100
PRDM2	chr1	14026604	14026606	cg09765994	6.5614	18.6329	1.87E-14	6.40E-12	0.0233	0.6696	0.0246	0.0293	0.0144
ADHFE1	chr8	67344587	67344589	cg18065361	7.5235	19.0847	1.17E-14	4.62E-12	0.0113	0.6664	0.0618	0.0139	0.0100
WHAMMP2 pseudogene	chr15	28983116	28983118	cg11794877	7.3934	29.8372	1.48E-18	8.22E-15	0.0124	0.6663	0.0177	0.0105	0.0100
TNFSF11	chr13	43148934	43148936	cg26224671	6.6724	13.8532	5.72E-12	4.43E-10	0.0208	0.6633	0.0520	0.0379	0.0235
CPLX2	chr5	175223981	175223983	cg07295964	6.8016	13.4048	1.06E-11	7.20E-10	0.0188	0.6621	0.0987	0.0529	0.0145
BVES	chr6	105584550	105584552	cg14159026	6.3043	19.9856	4.70E-15	2.41E-12	0.0270	0.6596	0.0835	0.0445	0.0184
H2AFY2	chr10	71813352	71813354	cg19726179	6.7892	17.4173	7.03E-14	1.65E-11	0.0187	0.6591	0.0607	0.0322	0.0134
T protein_coding	chr6	166582205	166582207	cg19675288	5.7797	15.9228	4.01E-13	5.85E-11	0.0403	0.6574	0.0837	0.0580	0.0219
SHD	chr19	4279440	4279442	cg26646370	4.9837	13.7830	6.29E-12	4.77E-10	0.0777	0.6495	0.0629	0.0835	0.0692
TRBJ2-1	chr7	142494913	142494915	cg21621906	7.3948	22.0045	6.92E-16	6.09E-13	0.0114	0.6479	0.0846	0.0111	0.0104
DRD4	chr11	637174	637176	cg12928379	5.5851	18.8822	1.44E-14	5.30E-12	0.0444	0.6459	0.0492	0.0361	0.0221
ZNF790	chr19	37329380	37329382	cg18876786	7.3732	20.8873	1.96E-15	1.28E-12	0.0114	0.6452	0.0185	0.0190	0.0116

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
TRPC4	chr13	38443949	38443951	cg19275632	6.1212	10.0497	1.96E-09	5.17E-08	0.0288	0.6447	0.0710	0.0368	0.0286
KCNQ3	chr8	133492475	133492477	cg27016990	7.4037	40.2134	3.31E-21	1.90E-16	0.0111	0.6435	0.0639	0.0271	0.0113
GALNT14	chr2	31360692	31360694	cg21583226	6.3703	19.6874	6.33E-15	2.99E-12	0.0236	0.6432	0.0544	0.0469	0.0135
TRBJ2-1	chr7	142494952	142494954	cg08430489	6.1580	17.9820	3.77E-14	1.05E-11	0.0276	0.6422	0.1000	0.0409	0.0217
EPB41L3	chr18	5629682	5629684	cg03718824	5.9428	15.8703	4.28E-13	6.14E-11	0.0320	0.6384	0.0433	0.0883	0.0371
TMEM178B	chr7	140773735	140773737	cg16633901	5.7002	19.6729	6.42E-15	3.02E-12	0.0387	0.6380	0.0804	0.0766	0.0398
EOMES	chr3	27763101	27763103	cg24434959	6.1404	20.5374	2.74E-15	1.64E-12	0.0272	0.6365	0.0278	0.0403	0.0228
CAMK2B	chr7	44364924	44364926	cg17035091	7.1596	28.6349	3.42E-18	1.47E-14	0.0127	0.6354	0.0132	0.0193	0.0128
GRID2	chr4	93226379	93226381	cg02468050	6.1239	13.6981	7.07E-12	5.22E-10	0.0274	0.6351	0.0751	0.0386	0.0211
TLL1	chr4	166794785	166794787	cg24521633	6.5363	15.1003	1.11E-12	1.27E-10	0.0200	0.6349	0.0420	0.0286	0.0146
RUNDC3B	chr7	87257766	87257768	cg15414833	5.5316	18.2767	2.74E-14	8.34E-12	0.0435	0.6344	0.0433	0.0521	0.0376
PRDM2	chr1	14026589	14026591	cg05346841	7.1979	22.6269	3.96E-16	4.10E-13	0.0123	0.6341	0.0208	0.0277	0.0100
N/A	chr7	1709949	1709951	cg10669265	6.3108	16.2049	2.86E-13	4.58E-11	0.0236	0.6335	0.0870	0.0384	0.0144
MPPED2	chr11	30607067	30607069	cg11855526	6.2535	25.9300	2.56E-17	6.00E-14	0.0245	0.6324	0.0494	0.0373	0.0153
FOXI3	chr2	88752321	88752323	cg23511613	7.4235	32.2323	3.08E-19	2.57E-15	0.0104	0.6322	0.0989	0.0125	0.0105
L3MBTL4	chr18	6414973	6414975	cg18556788	6.3332	13.5215	9.02E-12	6.35E-10	0.0229	0.6307	0.0513	0.0426	0.0172
ZNF665	chr19	53696648	53696650	cg01620580	6.7938	15.9956	3.67E-13	5.50E-11	0.0161	0.6291	0.0731	0.0154	0.0161
SIX2	chr2	45237688	45237690	cg02711647	5.5078	9.7187	3.50E-09	8.45E-08	0.0431	0.6288	0.0893	0.0186	0.0228
ZNF790	chr19	37341733	37341735	cg12033943	5.5624	16.3561	2.39E-13	4.01E-11	0.0401	0.6235	0.0227	0.0355	0.0389
GRIN3A	chr9	104501029	104501031	cg18794577	6.0871	15.9796	3.75E-13	5.58E-11	0.0261	0.6195	0.0599	0.0447	0.0320
ADARB2	chr10	1779834	1779836	cg02899206	5.6091	17.8687	4.26E-14	1.15E-11	0.0377	0.6191	0.0275	0.0456	0.0377
SNX32	chr11	65601331	65601333	cg11198128	7.3460	32.3319	2.89E-19	2.52E-15	0.0103	0.6191	0.0677	0.0100	0.0100
BEND5	chr1	49242518	49242520	cg16573178	6.3514	26.3634	1.83E-17	4.72E-14	0.0213	0.6188	0.0199	0.0295	0.0128
FGF14-IT1 lincRNA	chr13	103047052	103047054	cg16398329	7.2364	35.2351	4.98E-20	6.74E-16	0.0111	0.6182	0.0942	0.0141	0.0100
N/A	chr10	131769580	131769582	cg15344419	5.2108	20.9298	1.88E-15	1.25E-12	0.0519	0.6178	0.0818	0.0333	0.0230
B3GAT2	chr6	71666681	71666683	cg16556906	6.7127	23.1880	2.43E-16	2.87E-13	0.0158	0.6121	0.0858	0.0185	0.0156
BMP3	chr4	81951955	81951957	cg22403273	6.0836	13.4772	9.59E-12	6.65E-10	0.0251	0.6111	0.0170	0.0442	0.0184
ALX4	chr11	44330902	44330904	cg26365854	6.7108	20.8148	2.10E-15	1.34E-12	0.0158	0.6107	0.0915	0.0128	0.0100
CD1D	chr1	158150648	158150650	cg12124922	5.7274	17.0940	1.01E-13	2.14E-11	0.0328	0.6099	0.0273	0.0600	0.0211
ZNF790	chr19	37341869	37341871	cg03100040	5.1806	12.4490	4.20E-11	2.14E-09	0.0507	0.6087	0.0394	0.0585	0.0357

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
ZNF569	chr19	37957994	37957996	cg14142713	5.2517	23.4589	1.92E-16	2.47E-13	0.0475	0.6076	0.0559	0.0363	0.0506
EPDR1	chr7	37960872	37960874	cg10876076	4.8110	10.3249	1.22E-09	3.47E-08	0.0694	0.6074	0.0519	0.0749	0.0246
EPHA6	chr3	96533289	96533291	cg15093079	6.7418	21.4263	1.18E-15	9.09E-13	0.0149	0.6034	0.0918	0.0439	0.0157
ZNF71	chr19	57106537	57106539	cg22995684	6.0414	21.7878	8.43E-16	7.16E-13	0.0250	0.6030	0.0632	0.0298	0.0258
TSLP	chr5	110408996	110408998	cg24994173	6.9507	19.5733	7.10E-15	3.26E-12	0.0127	0.6013	0.0515	0.0312	0.0115
CNTNAP2	chr7	145813007	145813009	cg09571420	6.2109	19.4319	8.19E-15	3.58E-12	0.0217	0.6001	0.0950	0.0307	0.0107
L3MBTL4	chr18	6414957	6414959	cg22058122	6.1714	19.7903	5.71E-15	2.76E-12	0.0223	0.5999	0.0411	0.0381	0.0213
ZNF569	chr19	37957996	37957998	cg03884783	5.0340	23.7619	1.48E-16	2.08E-13	0.0546	0.5996	0.0610	0.0474	0.0596
AP000783.1	chr11	123301170	123301172	cg03401096	6.3640	29.2060	2.29E-18	1.08E-14	0.0192	0.5980	0.0116	0.0365	0.0100
ADARB2	chr10	1778843	1778845	cg07025949	5.8016	19.3253	9.13E-15	3.89E-12	0.0293	0.5978	0.0567	0.0527	0.0316
EPB41L3	chr18	5629179	5629181	cg07003920	5.2372	20.3938	3.15E-15	1.80E-12	0.0455	0.5970	0.0484	0.0633	0.0390
ZNF536	chr19	30718423	30718425	cg00382463	5.8127	15.5748	6.14E-13	8.05E-11	0.0288	0.5959	0.0350	0.0445	0.0204
BCL2	chr18	60985731	60985733	cg24740531	7.1697	20.1051	4.17E-15	2.23E-12	0.0105	0.5944	0.0106	0.0283	0.0107
ALX4	chr11	44331107	44331109	cg24335138	6.0367	16.3539	2.39E-13	4.01E-11	0.0241	0.5940	0.0832	0.0215	0.0196
EMX1	chr2	73151594	73151596	cg16781647	5.5478	18.5085	2.14E-14	7.04E-12	0.0349	0.5936	0.0600	0.0651	0.0146
RNF217	chr6	125284358	125284360	cg06993703	5.0805	16.5282	1.95E-13	3.48E-11	0.0496	0.5888	0.0923	0.0553	0.0379
LRRFIP1	chr2	238535909	238535911	cg15674193	5.6697	16.7889	1.44E-13	2.78E-11	0.0310	0.5886	0.0831	0.0320	0.0273
BEND5	chr1	49242512	49242514	cg11666087	6.1358	25.9571	2.50E-17	5.93E-14	0.0218	0.5883	0.0157	0.0155	0.0146
N/A	chr1	158119672	158119674	cg18873957	5.1725	11.3588	2.23E-10	8.36E-09	0.0455	0.5869	0.0575	0.0642	0.0786
ZNF736	chr7	63767840	63767842	cg23849078	7.0027	39.9533	3.79E-21	1.93E-16	0.0114	0.5855	0.0794	0.0126	0.0111
POU4F1	chr13	79177194	79177196	cg02279670	6.4530	23.5191	1.82E-16	2.38E-13	0.0169	0.5835	0.0875	0.0359	0.0113
EPB41L3	chr18	5544230	5544232	cg16924702	6.8475	20.3432	3.30E-15	1.87E-12	0.0126	0.5816	0.0320	0.0487	0.0119
RP4-555D20.4 lincRNA	chr3	44041024	44041026	cg07451080	6.1961	16.4152	2.23E-13	3.81E-11	0.0202	0.5816	0.0748	0.0420	0.0122
METTL24	chr6	110679106	110679108	cg10259748	4.8461	17.4090	7.10E-14	1.66E-11	0.0581	0.5815	0.0723	0.0545	0.0372
ZNF568	chr19	37407461	37407463	cg03443751	6.2568	16.0396	3.48E-13	5.28E-11	0.0192	0.5808	0.0546	0.0183	0.0159
N/A	chr1	164290544	164290546	cg24862510	6.1779	12.6454	3.14E-11	1.70E-09	0.0201	0.5776	0.0615	0.0407	0.0109
RUNDC3B	chr7	87258093	87258095	cg10600178	4.8995	14.0011	4.68E-12	3.79E-10	0.0543	0.5773	0.0284	0.0692	0.0243
MEGF8	chr19	42828155	42828157	cg11957331	5.7002	16.4698	2.09E-13	3.66E-11	0.0287	0.5769	0.0948	0.0558	0.0145
CALCA	chr11	14995200	14995202	cg22863523	6.6162	15.0898	1.13E-12	1.28E-10	0.0144	0.5754	0.0768	0.0238	0.0139
EGR2	chr10	64575441	64575443	cg07852757	5.1116	18.7759	1.61E-14	5.72E-12	0.0451	0.5753	0.0746	0.0863	0.0562

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
N/A	chr10	131770222	131770224	cg25446309	5.8829	17.1955	9.03E-14	1.97E-11	0.0247	0.5741	0.0456	0.0302	0.0262
EPB41L3	chr18	5544098	5544100	cg16622495	6.8115	24.3432	9.13E-17	1.52E-13	0.0124	0.5732	0.0355	0.0387	0.0103
RCSD1	chr1	167599718	167599720	cg18800085	7.0311	33.3011	1.58E-19	1.51E-15	0.0106	0.5727	0.0725	0.0222	0.0100
GABRA4	chr4	46995820	46995822	cg16976370	5.8633	15.3667	7.95E-13	9.79E-11	0.0247	0.5708	0.0409	0.0306	0.0215
BMP3	chr4	81951950	81951952	cg20642710	6.3442	15.1842	1.00E-12	1.16E-10	0.0171	0.5690	0.0175	0.0249	0.0171
ZNF582	chr19	56905093	56905095	cg25267765	6.6918	18.0420	3.53E-14	1.01E-11	0.0133	0.5688	0.0424	0.0224	0.0100
N/A	chr4	180979630	180979632	cg12950911	5.1801	18.8264	1.53E-14	5.50E-12	0.0413	0.5684	0.0640	0.0561	0.0251
FBN1	chr15	48938346	48938348	cg10480343	6.0087	20.3711	3.22E-15	1.83E-12	0.0218	0.5680	0.0986	0.0321	0.0189
PDE3A	chr12	20521462	20521464	cg17535647	4.9003	16.4560	2.12E-13	3.70E-11	0.0517	0.5677	0.0707	0.0927	0.0207
CD1D	chr1	158150796	158150798	cg04360049	5.1061	16.8909	1.28E-13	2.54E-11	0.0434	0.5666	0.0692	0.0636	0.0423
ZNF470	chr19	57078441	57078443	cg03153658	6.1502	18.4341	2.31E-14	7.37E-12	0.0195	0.5657	0.0328	0.0212	0.0112
WNT5A	chr3	55521350	55521352	cg22615158	6.9736	21.3848	1.22E-15	9.37E-13	0.0107	0.5652	0.0803	0.0100	0.0105
ATRNL1	chr10	116854027	116854029	cg24252925	5.5557	18.6848	1.77E-14	6.16E-12	0.0303	0.5648	0.0800	0.0784	0.0229
ZNF470	chr19	57078764	57078766	cg09265529	6.6193	39.0290	6.12E-21	2.35E-16	0.0137	0.5630	0.0792	0.0283	0.0143
ZNF665	chr19	53696641	53696643	cg07686635	4.9633	16.5287	1.95E-13	3.48E-11	0.0478	0.5624	0.0796	0.0509	0.0350
KCNN2	chr5	113697329	113697331	cg26063563	6.4189	25.1290	4.81E-17	9.42E-14	0.0157	0.5618	0.0984	0.0151	0.0116
DYDC1	chr10	82117088	82117090	cg03701427	5.8448	14.5511	2.25E-12	2.16E-10	0.0240	0.5617	0.0892	0.0509	0.0169
ZNF470	chr19	57078779	57078781	cg24334111	6.2114	32.7921	2.17E-19	2.03E-15	0.0183	0.5612	0.0618	0.0337	0.0104
COL4A4	chr2	228029516	228029518	cg22388982	4.9251	16.9820	1.15E-13	2.35E-11	0.0489	0.5610	0.0956	0.0603	0.0550
ZNF528	chr19	52900973	52900975	cg27258025	5.3675	20.0356	4.47E-15	2.34E-12	0.0338	0.5569	0.0891	0.0568	0.0235
FAM181B	chr11	82444797	82444799	cg09617579	4.9178	25.0101	5.30E-17	1.01E-13	0.0482	0.5568	0.0997	0.0802	0.0223
TMEM178B	chr2	39893077	39893079	cg19935171	6.4207	17.2647	8.35E-14	1.87E-11	0.0153	0.5560	0.0661	0.0540	0.0163
SLC8A1	chr2	40679518	40679520	cg03307465	5.2186	18.1299	3.21E-14	9.37E-12	0.0375	0.5546	0.0576	0.0468	0.0153
ZNF569 / ZNF570	chr19	37958345	37958347	cg08630279	5.7623	37.9997	1.06E-20	2.87E-16	0.0246	0.5531	0.0259	0.0251	0.0244
PLXNA4	chr7	132261417	132261419	cg05950570	6.3815	17.4361	6.89E-14	1.63E-11	0.0154	0.5508	0.0374	0.0361	0.0127
CPE	chr4	166300251	166300253	cg16534103	5.1693	18.3062	2.65E-14	8.14E-12	0.0383	0.5508	0.0400	0.0338	0.0284
ZNF790	chr19	37329273	37329275	cg22379207	6.4450	19.9861	4.70E-15	2.41E-12	0.0146	0.5496	0.0200	0.0129	0.0208
NAT16	chr7	100823376	100823378	cg14562712	4.6498	19.6985	6.26E-15	2.97E-12	0.0577	0.5483	0.0884	0.0738	0.0402
TLL1	chr4	166794970	166794972	cg19898128	6.4722	22.5648	4.19E-16	4.28E-13	0.0143	0.5480	0.0293	0.0358	0.0106
RP11-433J8.1 lincRNA	chr14	97059191	97059193	cg24034005	6.2511	21.2393	1.40E-15	1.02E-12	0.0167	0.5479	0.0549	0.0352	0.0160

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
ZNF331	chr19	54024322	54024324	cg15205507	5.9858	19.9968	4.65E-15	2.40E-12	0.0202	0.5467	0.0470	0.0217	0.0166
BMP3	chr4	81951952	81951954	cg26105156	5.5803	18.0396	3.54E-14	1.01E-11	0.0274	0.5465	0.0200	0.0496	0.0100
HS3ST4	chr16	25703829	25703831	cg26734696	6.1888	16.3548	2.39E-13	4.01E-11	0.0174	0.5465	0.0296	0.0187	0.0136
ZNF790	chr19	37329447	37329449	cg06466031	5.0846	14.3001	3.14E-12	2.78E-10	0.0398	0.5445	0.0768	0.0472	0.0420
TLL1	chr4	166795523	166795525	cg10756127	5.3673	15.6897	5.33E-13	7.25E-11	0.0319	0.5444	0.0584	0.0647	0.0338
N/A	chr11	132934553	132934555	cg02224864	4.8669	15.7233	5.12E-13	7.03E-11	0.0472	0.5441	0.0979	0.0570	0.0287
ZNF71	chr19	57106851	57106853	cg04181321	5.8695	15.7170	5.16E-13	7.07E-11	0.0218	0.5441	0.0494	0.0211	0.0137
AP000783.1	chr11	123300866	123300868	cg19419279	5.3960	18.4616	2.25E-14	7.25E-12	0.0309	0.5425	0.0319	0.0576	0.0297
SLC30A2	chr1	26372894	26372896	cg00703260	5.6972	18.0195	3.62E-14	1.02E-11	0.0246	0.5423	0.0582	0.0483	0.0179
NPAS3	chr14	33403109	33403111	cg18945335	5.9024	13.6155	7.92E-12	5.73E-10	0.0211	0.5422	0.0982	0.0395	0.0191
N/A	chr11	123229179	123229181	cg10096161	4.9694	21.8788	7.76E-16	6.68E-13	0.0429	0.5411	0.0597	0.0703	0.0302
RFTN1	chr3	16554465	16554467	cg04066019	6.2008	15.9765	3.76E-13	5.60E-11	0.0168	0.5408	0.0678	0.0242	0.0119
ADARB2	chr10	1780002	1780004	cg23684973	4.4698	21.3153	1.31E-15	9.70E-13	0.0648	0.5407	0.0898	0.0955	0.0784
CD1D	chr1	158151362	158151364	cg05338433	6.0783	19.1650	1.08E-14	4.36E-12	0.0183	0.5389	0.0801	0.0433	0.0165
N/A	chr5	159399505	159399507	cg03825010	5.5197	15.7795	4.78E-13	6.67E-11	0.0277	0.5385	0.0997	0.0265	0.0220
ZNF569 / ZNF570	chr19	37960106	37960108	cg17778441	5.0898	24.0350	1.18E-16	1.79E-13	0.0383	0.5371	0.0352	0.0301	0.0337
ZNF141	chr4	330693	330695	cg11510060	5.8515	20.3210	3.38E-15	1.89E-12	0.0214	0.5369	0.0573	0.0387	0.0243
ZNF570	chr19	37960413	37960415	cg11595155	6.1399	13.6933	7.11E-12	5.25E-10	0.0173	0.5369	0.0129	0.0100	0.0145
PLXNC1	chr12	94543448	94543450	cg13565656	4.5538	18.2670	2.77E-14	8.40E-12	0.0582	0.5338	0.0663	0.0604	0.0128
NDRG4	chr16	58498150	58498152	cg00984694	5.7637	22.4872	4.48E-16	4.51E-13	0.0224	0.5324	0.0476	0.0310	0.0136
UBE2QL1	chr5	6449089	6449091	cg06997381	6.8403	33.4199	1.47E-19	1.47E-15	0.0101	0.5301	0.0670	0.0100	0.0100
CTD-2162K18.5 antisense	chr19	37288704	37288706	cg00578154	6.7525	19.0238	1.25E-14	4.81E-12	0.0108	0.5297	0.0639	0.0171	0.0108
GCM2	chr6	10882239	10882241	cg14250833	5.2099	10.7123	6.38E-10	2.01E-08	0.0337	0.5295	0.0974	0.0585	0.0270
EPB41L3	chr18	5629370	5629372	cg24496223	5.9115	23.1556	2.49E-16	2.93E-13	0.0198	0.5289	0.0496	0.0565	0.0232
L3MBTL4	chr18	6414975	6414977	cg14352983	5.3618	11.1962	2.90E-10	1.04E-08	0.0295	0.5261	0.0687	0.0623	0.0232
ZNF528	chr19	52901123	52901125	cg03091337	5.2010	27.9030	5.79E-18	2.17E-14	0.0331	0.5244	0.0993	0.0396	0.0327
ASIC2	chr17	31618408	31618410	cg02984614	4.2779	13.1360	1.55E-11	9.71E-10	0.0697	0.5227	0.0584	0.0779	0.0134
LRRC10B	chr11	61277016	61277018	cg06341513	6.6246	19.4110	8.37E-15	3.64E-12	0.0113	0.5183	0.0485	0.0343	0.0100
N/A	chr1	158083474	158083476	cg13930105	6.6645	35.5015	4.27E-20	6.63E-16	0.0110	0.5183	0.0115	0.0151	0.0106
FER1L4	chr20	34189410	34189412	cg13544006	4.8011	17.2086	8.90E-14	1.96E-11	0.0439	0.5175	0.0447	0.0878	0.0480

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
pseudogene													
BICC1	chr10	60272753	60272755	cg07857251	6.2585	27.1340	1.02E-17	3.13E-14	0.0146	0.5175	0.0772	0.0232	0.0140
ZNF793	chr19	37997293	37997295	cg11623861	6.5465	27.1097	1.04E-17	3.14E-14	0.0119	0.5173	0.0852	0.0224	0.0100
SLC22A31	chr16	89267825	89267827	cg00588720	5.5521	19.0896	1.16E-14	4.61E-12	0.0245	0.5166	0.0537	0.0368	0.0229
BEND5	chr1	49242756	49242758	cg03603214	6.1778	33.6681	1.26E-19	1.32E-15	0.0154	0.5152	0.0102	0.0116	0.0100
N/A	chr1	180204220	180204222	cg05719164	5.1254	11.3222	2.37E-10	8.79E-09	0.0335	0.5143	0.0787	0.0689	0.0229
DYDC1/DYDC2	chr10	82116364	82116366	cg11854806	6.2211	23.9412	1.28E-16	1.88E-13	0.0148	0.5139	0.0731	0.0228	0.0109
AJAP1	chr1	4716536	4716538	cg02483484	6.2899	28.0591	5.17E-18	2.02E-14	0.0139	0.5099	0.0766	0.0273	0.0120
FJX1	chr11	35641504	35641506	cg02370395	4.2562	16.5442	1.91E-13	3.44E-11	0.0657	0.5078	0.0843	0.0580	0.0651
EPB41L3	chr18	5544236	5544238	cg22335490	6.2575	17.7290	4.97E-14	1.28E-11	0.0140	0.5065	0.0190	0.0304	0.0111
N/A	chr11	132934138	132934140	cg00232735	5.5297	17.6758	5.27E-14	1.34E-11	0.0238	0.5054	0.0947	0.0702	0.0155
GALNT14	chr2	31360816	31360818	cg03279535	4.7489	20.1652	3.93E-15	2.12E-12	0.0431	0.5047	0.0880	0.0640	0.0415
N/A	chr17	47301740	47301742	cg15262242	4.7881	18.0917	3.34E-14	9.65E-12	0.0414	0.5026	0.0334	0.0370	0.0268
WSCD1	chr17	5974077	5974079	cg27112897	5.8788	21.0772	1.63E-15	1.13E-12	0.0181	0.5025	0.0702	0.0276	0.0109
BMP3	chr4	81951471	81951473	cg02621694	4.4591	12.8704	2.27E-11	1.31E-09	0.0537	0.5016	0.0454	0.0918	0.0662
BMP3	chr4	81952329	81952331	cg20276585	5.9374	28.1406	4.88E-18	1.93E-14	0.0171	0.4983	0.0163	0.0326	0.0109
CLSTN2	chr3	139653544	139653546	cg01755562	4.7459	9.0599	1.15E-08	2.35E-07	0.0419	0.4978	0.0865	0.0314	0.0124
UNC5A	chr5	176237220	176237222	cg13928709	4.5310	12.5583	3.57E-11	1.89E-09	0.0497	0.4978	0.0716	0.0906	0.0151
CPE	chr4	166300016	166300018	cg03465206	6.4812	20.0086	4.59E-15	2.38E-12	0.0114	0.4966	0.0222	0.0146	0.0108
HOPX	chr4	57522144	57522146	cg09853371	4.4606	19.9234	5.00E-15	2.52E-12	0.0524	0.4965	0.0426	0.0980	0.0453
ZNF569 / ZNF570	chr19	37958452	37958454	cg18244915	5.2442	25.1673	4.67E-17	9.26E-14	0.0283	0.4963	0.0217	0.0230	0.0285
ST8SIA1	chr12	22487667	22487669	cg12253830	5.6819	14.7149	1.82E-12	1.83E-10	0.0201	0.4929	0.0846	0.0348	0.0122
WSCD1	chr17	5974065	5974067	cg25578609	6.0146	32.3166	2.92E-19	2.52E-15	0.0156	0.4908	0.0618	0.0246	0.0144
C12orf39	chr12	21680507	21680509	cg10437806	5.0995	23.2249	2.35E-16	2.81E-13	0.0305	0.4886	0.0998	0.0435	0.0181
PRKG2	chr4	82136838	82136840	cg13798300	4.9784	10.1719	1.59E-09	4.32E-08	0.0332	0.4864	0.0313	0.0326	0.0100
N/A	chr8	143591595	143591597	cg07139330	6.1852	14.1885	3.64E-12	3.12E-10	0.0135	0.4849	0.0868	0.0134	0.0141
ZNF569 / ZNF570	chr19	37958443	37958445	cg25451874	4.9866	22.7739	3.48E-16	3.75E-13	0.0326	0.4842	0.0345	0.0351	0.0362
AP000783.1	chr11	123301673	123301675	cg05289253	4.9930	28.6174	3.47E-18	1.48E-14	0.0324	0.4839	0.0236	0.0733	0.0243
ALK	chr2	30144578	30144580	cg14163665	5.1610	10.6445	7.14E-10	2.20E-08	0.0285	0.4832	0.0787	0.0361	0.0136
AP000783.1	chr11	123301489	123301491	cg10904699	4.3111	25.3078	4.17E-17	8.61E-14	0.0555	0.4829	0.0417	0.0700	0.0264
TCF24	chr8	67874205	67874207	cg26618965	5.2840	17.8850	4.19E-14	1.14E-11	0.0257	0.4811	0.0933	0.0284	0.0129

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
ZNF569 / ZNF570	chr19	37960541	37960543	cg02385791	4.7901	14.4822	2.47E-12	2.31E-10	0.0371	0.4789	0.0338	0.0346	0.0396
AC114730.5 antisense	chr2	242743213	242743215	cg03462699	5.0963	16.7562	1.49E-13	2.85E-11	0.0293	0.4788	0.0273	0.0990	0.0194
N/A	chr4	4867101	4867103	cg12378888	5.2164	10.6356	7.24E-10	2.23E-08	0.0267	0.4784	0.0797	0.0500	0.0106
MATK	chr19	3786124	3786126	cg13535593	4.6430	13.0909	1.65E-11	1.02E-09	0.0410	0.4754	0.0982	0.0505	0.0247
ZEB2	chr2	145274974	145274976	cg11540007	6.4930	33.8973	1.10E-19	1.23E-15	0.0104	0.4753	0.0944	0.0283	0.0100
EBF2	chr8	25901421	25901423	cg16749235	4.7995	11.4351	1.98E-10	7.57E-09	0.0359	0.4734	0.0876	0.0465	0.0103
ZNF43	chr19	22018952	22018954	cg13604887	4.7725	23.2484	2.30E-16	2.79E-13	0.0360	0.4690	0.0851	0.0456	0.0502
DMRT1	chr9	841558	841560	cg13762060	5.9164	18.1889	3.01E-14	8.96E-12	0.0153	0.4690	0.0964	0.0201	0.0173
CALY	chr10	135150427	135150429	cg14614314	5.5752	15.8077	4.61E-13	6.49E-11	0.0195	0.4677	0.0193	0.0591	0.0174
TMEM200C	chr18	5896266	5896268	cg23841467	3.8430	10.8061	5.46E-10	1.77E-08	0.0745	0.4616	0.0926	0.0878	0.0698
ST8SIA1	chr12	22487657	22487659	cg16970851	5.6456	15.4951	6.78E-13	8.67E-11	0.0178	0.4585	0.0897	0.0287	0.0100
ANKRD24	chr19	4198057	4198059	cg14339043	5.1927	15.5132	6.63E-13	8.52E-11	0.0248	0.4568	0.0428	0.0433	0.0169
N/A	chr18	77558845	77558847	cg13135595	5.0720	15.0915	1.12E-12	1.28E-10	0.0271	0.4567	0.0877	0.0456	0.0108
GPR156	chr3	120004185	120004187	cg15602740	5.0729	6.1461	4.43E-06	3.98E-05	0.0270	0.4556	0.0723	0.0428	0.0168
ZNF610	chr19	52839935	52839937	cg09383947	4.1594	18.5761	1.99E-14	6.69E-12	0.0532	0.4479	0.0910	0.0474	0.0302
ZNF536	chr19	30719148	30719150	cg00757082	4.6704	13.8622	5.65E-12	4.39E-10	0.0351	0.4457	0.0736	0.0537	0.0420
PRKG2	chr4	82136115	82136117	cg01508023	4.2530	14.5317	2.31E-12	2.20E-10	0.0487	0.4450	0.0649	0.0827	0.0258
LRRC10B	chr11	61276446	61276448	cg25923577	5.4033	8.6520	2.48E-08	4.51E-07	0.0201	0.4448	0.0796	0.0402	0.0143
NELL1	chr11	20691428	20691430	cg01563031	6.0000	18.6232	1.89E-14	6.44E-12	0.0128	0.4416	0.0827	0.0164	0.0107
HAND1	chr5	153858821	153858823	cg15376615	5.3663	12.7541	2.68E-11	1.50E-09	0.0203	0.4409	0.0774	0.0178	0.0100
HCN1	chr5	45696464	45696466	cg19697475	5.1665	13.9734	4.86E-12	3.90E-10	0.0231	0.4362	0.0564	0.0562	0.0112
CBLN1	chr16	49316882	49316884	cg07571509	5.1657	10.2443	1.40E-09	3.90E-08	0.0229	0.4335	0.0512	0.0475	0.0391
GREB1L	chr18	18821912	18821914	cg06711831	3.6164	5.3134	2.96E-05	0.000202799	0.0787	0.4330	0.0884	0.0908	0.0287
TRHDE	chr12	72666263	72666265	cg02511156	5.3378	15.5085	6.67E-13	8.56E-11	0.0199	0.4309	0.0984	0.0341	0.0123
EMX2	chr10	119304582	119304584	cg03339065	4.3591	6.0540	5.45E-06	4.76E-05	0.0409	0.4256	0.0855	0.0539	0.0387
TLL1	chr4	166795259	166795261	cg12836011	4.4886	15.7833	4.75E-13	6.65E-11	0.0368	0.4249	0.0591	0.0726	0.0302
CTNNA2	chr2	79739939	79739941	cg08047376	4.4171	8.3993	4.03E-08	6.84E-07	0.0381	0.4203	0.0464	0.0351	0.0135
ZNF528	chr19	52900926	52900928	cg20725941	5.0897	11.1858	2.95E-10	1.06E-08	0.0225	0.4170	0.0715	0.0329	0.0204
N/A	chr11	132864164	132864166	cg27312388	3.8247	5.8726	8.20E-06	6.75E-05	0.0603	0.4159	0.0810	0.0512	0.0268
GPM6A	chr4	176922778	176922780	cg08576864	3.9045	7.9689	9.38E-08	1.42E-06	0.0553	0.4120	0.0666	0.0690	0.0445

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
CTNNA2	chr2	79740321	79740323	cg01546568	3.4034	6.3201	3.01E-06	2.85E-05	0.0855	0.4119	0.0691	0.0460	0.0521
TMEM200C	chr18	5895538	5895540	cg20523393	6.0882	17.3377	7.69E-14	1.75E-11	0.0106	0.4101	0.0812	0.0117	0.0122
N/A	chr13	112712474	112712476	cg03276408	3.4408	12.2201	5.91E-11	2.82E-09	0.0817	0.4097	0.0861	0.0565	0.0137
ZNF331	chr19	54023998	54024000	cg11332363	3.8060	9.6065	4.27E-09	1.00E-07	0.0590	0.4081	0.0434	0.0904	0.0452
N/A	chr6	137819266	137819268	cg19003797	4.8572	10.8726	4.90E-10	1.61E-08	0.0257	0.4080	0.0453	0.0437	0.0100
ZNF829	chr19	37407059	37407061	cg02370417	5.8801	24.4497	8.36E-17	1.43E-13	0.0120	0.4052	0.0848	0.0204	0.0140
NXPH3	chr17	47653305	47653307	cg15167646	4.0576	5.5329	1.78E-05	0.00013133	0.0469	0.4034	0.0370	0.0568	0.0259
RBM11	chr21	15588607	15588609	cg16492707	5.7822	11.1165	3.29E-10	1.16E-08	0.0127	0.4024	0.0218	0.0100	0.0119
GALNTL6	chr4	172734506	172734508	cg14307443	5.1480	11.8680	1.01E-10	4.36E-09	0.0196	0.3953	0.0947	0.0220	0.0114
ALX1	chr12	85674693	85674695	cg00765312	3.5604	4.2547	0.000359866	0.001702469	0.0682	0.3952	0.0588	0.0931	0.0152
N/A	chr19	48918115	48918117	cg15779837	4.7671	9.2429	8.23E-09	1.75E-07	0.0258	0.3935	0.0981	0.0294	0.0147
CELF4	chr18	35147078	35147080	cg15179725	5.7606	17.1063	9.99E-14	2.12E-11	0.0117	0.3799	0.0930	0.0118	0.0126
FOXB1	chr15	60298042	60298044	cg04861263	5.7211	14.5027	2.40E-12	2.26E-10	0.0121	0.3796	0.0523	0.0163	0.0142
TENM3	chr4	183066336	183066338	cg10603183	4.5886	8.1505	6.55E-08	1.04E-06	0.0274	0.3768	0.0467	0.0420	0.0342
SPATA32	chr17	43339588	43339590	cg24542751	3.6793	9.7360	3.39E-09	8.23E-08	0.0551	0.3724	0.0548	0.0549	0.0157
HRH2	chr5	175084742	175084744	cg17483297	4.0517	9.3242	7.10E-09	1.55E-07	0.0398	0.3676	0.0541	0.0462	0.0440
EMX2	chr10	119304585	119304587	cg24188415	4.9120	6.8287	9.94E-07	1.10E-05	0.0198	0.3582	0.0590	0.0445	0.0180
SYT6	chr1	114696461	114696463	cg09177131	5.0757	14.7079	1.84E-12	1.84E-10	0.0173	0.3555	0.0982	0.0263	0.0104
HOXC8	chr12	54403177	54403179	cg02344911	3.9058	7.2538	4.04E-07	5.03E-06	0.0395	0.3416	0.0775	0.0700	0.0434
EDARADD	chr1	236558873	236558875	cg23020486	5.6803	7.7715	1.39E-07	2.00E-06	0.0101	0.3344	0.0163	0.0105	0.0100
RP4-555D20.3 lincRNA	chr3	44038249	44038251	cg14560001	3.6444	6.8465	9.57E-07	1.06E-05	0.0441	0.3218	0.0614	0.0488	0.0186

Hypermethylated probes for differentiation of intervention requiring disease from normal and non-dysplastic disease

165 probes (all shown), hypermethylated in HGD and EAC in comparison to non-dysplastic BE were identified using stringent cut-off criteria (minimal methylation in normal and control tissues, as well as normal peripheral blood). This is the most clinically relevant comparison, differentiating intervention-requiring disease from normal and non-dysplastic disease. Gencode v19 used for annotation[204], gene name protein-coding when not specified. logfc: log fold change, tstat: t-statistic, pval: p-value, adj_pval: adjusted p-value, β_{norm}: average methylation in normal squamous epitheilium, β_{base}: average methylation in baseline samples, here,

Appendix 7: Hypermethylation biomarker probes for BE and EAC

non-dysplastic BE epithelium, Δβ: difference in average methylation between comparison groups, here, between HGD-EAC and BE, β_{duo}: average methylation in duodenal epithelium, β_{stom}: average methylation in proximal stomach epithelium, β_{blood}: average methylation in peripheral blood from disease-free patients. Listed in order of decreasing Δβ. Entries in grey correspond to target region genes validated by targeted amplicon sequencing. Note that region selection for validation included criteria pertaining to identification of aberrant methylation across groups of probes (within 300bp, at least 2 differentially methylated probes required) as well as region suitability for BSP assay design. Hence some of the top aberrantly hypermethylated individual probes listed here were not part of regions selected for validation.

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{norm}	β_{base}	Δβ	β_{duo}	β_{stom}	β_{blood}
DCHS1	chr11	6677086	6677088	cg21012362	6.5492	11.4311	1.99E-10	7.61E-09	0.0197	0.0169	0.5998	0.0133	0.0227	0.0192
GNG4	chr1	235812839	235812841	cg06714284	7.1102	13.7287	6.78E-12	5.05E-10	0.0108	0.0108	0.5902	0.0100	0.0123	0.0100
DCBLD2	chr3	98620661	98620663	cg27352765	6.2823	8.4444	3.69E-08	6.35E-07	0.0194	0.0180	0.5702	0.0144	0.0168	0.0194
KCNB2	chr8	73449523	73449525	cg15774465	6.3478	10.6222	7.41E-10	2.27E-08	0.0176	0.0162	0.5564	0.0621	0.0274	0.0108
GBGT1	chr9	136039124	136039126	cg14472025	5.2494	13.0458	1.76E-11	1.07E-09	0.0359	0.0353	0.5466	0.0246	0.0384	0.0237
KCNB2	chr8	73449518	73449520	cg18555069	6.8205	12.7203	2.82E-11	1.56E-09	0.0108	0.0106	0.5376	0.0765	0.0113	0.0102
SH3RF3	chr2	109745647	109745649	cg16117910	5.5862	9.3091	7.30E-09	1.58E-07	0.0268	0.0259	0.5351	0.0211	0.0380	0.0162
PCLO	chr7	82792344	82792346	cg15478390	6.0999	12.8736	2.26E-11	1.31E-09	0.0185	0.0169	0.5247	0.0307	0.0144	0.0194
SH3RF3	chr2	109745613	109745615	cg22959667	6.1606	11.1374	3.18E-10	1.13E-08	0.0167	0.0161	0.5232	0.0124	0.0193	0.0103
FADS1	chr11	61583955	61583957	cg23992449	6.4606	10.6539	7.03E-10	2.18E-08	0.0128	0.0124	0.5127	0.0109	0.0143	0.0103
QPCT	chr2	37571728	37571730	cg15162392	4.8835	6.3856	2.61E-06	2.52E-05	0.0471	0.0401	0.5124	0.0465	0.0637	0.0373
N/A	chr5	36690436	36690438	cg05542957	6.6452	17.8258	4.47E-14	1.19E-11	0.0109	0.0108	0.5116	0.0564	0.0138	0.0155
N/A	chr7	127807499	127807501	cg25402610	6.3224	10.6097	7.56E-10	2.31E-08	0.0157	0.0136	0.5108	0.0105	0.0176	0.0100
FADS1	chr11	61584111	61584113	cg27173322	6.5875	10.1053	1.78E-09	4.76E-08	0.0114	0.0112	0.5105	0.0150	0.0100	0.0100
PCP4L1	chr1	161228495	161228497	cg27223727	6.1377	10.7388	6.10E-10	1.93E-08	0.0157	0.0147	0.4976	0.0139	0.0180	0.0119
GRASP	chr12	52400662	52400664	cg22196952	6.5829	12.9052	2.16E-11	1.26E-09	0.0108	0.0106	0.4972	0.0124	0.0121	0.0100
ENO4	chr10	118609166	118609168	cg15461105	5.0526	13.2572	1.31E-11	8.49E-10	0.0330	0.0324	0.4943	0.0233	0.0445	0.0259
FADS1	chr11	61583570	61583572	cg13475388	6.0038	8.4868	3.40E-08	5.92E-07	0.0168	0.0157	0.4900	0.0136	0.0111	0.0119
DNM3	chr1	171811476	171811478	cg20800956	5.9419	10.3139	1.24E-09	3.53E-08	0.0175	0.0160	0.4845	0.0618	0.0197	0.0102
HLX	chr1	221053512	221053514	cg15356516	4.5176	7.6909	1.64E-07	2.31E-06	0.0472	0.0443	0.4708	0.0344	0.0512	0.0214
C17orf64	chr17	58498976	58498978	cg09695735	4.7841	5.5127	1.87E-05	0.000136659	0.0423	0.0353	0.4665	0.0409	0.0100	0.0100
GRASP	chr12	52400666	52400668	cg09101796	6.1815	10.9734	4.15E-10	1.40E-08	0.0132	0.0123	0.4630	0.0161	0.0117	0.0100

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{norm}	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
EPHX3	chr19	15343344	15343346	cg17399362	3.7437	7.5042	2.40E-07	3.21E-06	0.0851	0.0811	0.4606	0.0725	0.0786	0.0721
ARHGAP22	chr10	49732002	49732004	cg11790857	4.9357	9.5057	5.11E-09	1.17E-07	0.0325	0.0303	0.4587	0.0314	0.0435	0.0233
MOXD1	chr6	132722314	132722316	cg16478774	6.1534	15.7335	5.05E-13	6.96E-11	0.0133	0.0123	0.4582	0.0446	0.0119	0.0100
RTN1	chr14	60337476	60337478	cg19556814	5.2710	6.9915	7.02E-07	8.13E-06	0.0241	0.0234	0.4568	0.0218	0.0362	0.0107
N/A	chr11	123947156	123947158	cg11633126	5.7072	10.9566	4.27E-10	1.44E-08	0.0186	0.0168	0.4545	0.0437	0.0291	0.0198
SLC16A14	chr2	230933066	230933068	cg26338195	4.8369	8.1124	7.06E-08	1.11E-06	0.0334	0.0317	0.4519	0.0266	0.0348	0.0196
DCBLD2	chr3	98620815	98620817	cg17198587	4.1042	8.1065	7.14E-08	1.12E-06	0.0569	0.0554	0.4467	0.0489	0.0523	0.0447
ZNF365	chr10	64133895	64133897	cg08984023	4.0808	12.3306	5.01E-11	2.47E-09	0.0587	0.0563	0.4462	0.0815	0.0489	0.0456
DCBLD2	chr3	98620856	98620858	cg02464093	5.8339	7.5392	2.24E-07	3.02E-06	0.0156	0.0147	0.4449	0.0103	0.0207	0.0118
BOC	chr3	112930674	112930676	cg27085904	4.4421	6.5456	1.84E-06	1.86E-05	0.0434	0.0416	0.4436	0.0550	0.0544	0.0339
N/A	chr10	102587109	102587111	cg26439963	4.1981	14.8407	1.55E-12	1.62E-10	0.0513	0.0504	0.4433	0.0349	0.0540	0.0291
EPHX3	chr19	15342748	15342750	cg03562044	5.4184	10.8754	4.87E-10	1.60E-08	0.0211	0.0194	0.4394	0.0521	0.0242	0.0119
ELOVL5	chr6	53212617	53212619	cg14378848	5.3004	10.0967	1.81E-09	4.82E-08	0.0218	0.0211	0.4383	0.0186	0.0264	0.0259
N/A	chr17	45867660	45867662	cg16624787	4.9431	10.5956	7.74E-10	2.36E-08	0.0282	0.0274	0.4367	0.0303	0.0195	0.0248
DTX3	chr12	57998655	57998657	cg19217692	3.8377	12.2221	5.89E-11	2.81E-09	0.0665	0.0657	0.4358	0.0631	0.0501	0.0548
RTN1	chr14	60337473	60337475	cg04510871	4.1255	6.9102	8.35E-07	9.44E-06	0.0515	0.0507	0.4316	0.0413	0.0495	0.0189
DTX3	chr12	57998580	57998582	cg06596054	3.7651	10.7294	6.20E-10	1.96E-08	0.0681	0.0665	0.4256	0.0529	0.0731	0.0889
RRAGD	chr6	90121835	90121837	cg11571263	4.0899	13.3066	1.22E-11	8.05E-10	0.0516	0.0506	0.4251	0.0451	0.0523	0.0469
NEUROG2	chr4	113437229	113437231	cg26708817	4.9705	10.1637	1.61E-09	4.36E-08	0.0259	0.0252	0.4228	0.0257	0.0281	0.0194
GBGT1	chr9	136039280	136039282	cg21244880	5.8679	14.1114	4.03E-12	3.39E-10	0.0134	0.0130	0.4220	0.0113	0.0165	0.0101
N/A	chr5	36690601	36690603	cg18711394	4.2676	14.5817	2.17E-12	2.10E-10	0.0435	0.0431	0.4214	0.0758	0.0313	0.0305
RP11-266L9.5 lincRNA	chr16	25078094	25078096	cg04153495	4.3415	7.1323	5.21E-07	6.27E-06	0.0403	0.0396	0.4160	0.0289	0.0445	0.0368
QPCT	chr2	37572288	37572290	cg17304496	3.6358	5.9972	6.19E-06	5.31E-05	0.0706	0.0701	0.4138	0.0576	0.0684	0.0622
N/A	chr10	102586998	102587000	cg25531836	3.8943	14.5757	2.18E-12	2.11E-10	0.0567	0.0560	0.4128	0.0600	0.0564	0.0658
TNFRSF9	chr1	8002527	8002529	cg18859763	4.7807	14.4678	2.51E-12	2.34E-10	0.0289	0.0273	0.4084	0.0534	0.0381	0.0249
RRAGD	chr6	90121832	90121834	cg21051989	3.5041	10.6344	7.26E-10	2.24E-08	0.0799	0.0763	0.4075	0.0352	0.0838	0.0465
ZNF365	chr10	64133887	64133889	cg01137532	4.2486	14.4466	2.58E-12	2.40E-10	0.0423	0.0407	0.4059	0.0884	0.0409	0.0477
CLDN10	chr13	96204859	96204861	cg16275739	4.4569	10.1397	1.68E-09	4.53E-08	0.0362	0.0345	0.4052	0.0770	0.0328	0.0256
NPAS3	chr14	33408947	33408949	cg04900565	4.0206	7.4394	2.75E-07	3.61E-06	0.0502	0.0484	0.4038	0.0432	0.0673	0.0282
SPOCK1	chr5	136835106	136835108	cg12420858	6.1255	12.2557	5.60E-11	2.70E-09	0.0100	0.0100	0.4036	0.0269	0.0100	0.0100

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{norm}	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
RLTPR	chr16	67678845	67678847	cg00846114	5.1760	7.9173	1.04E-07	1.55E-06	0.0223	0.0198	0.4028	0.0103	0.0258	0.0102
N/A	chr7	6566323	6566325	cg02266348	4.2644	10.7254	6.24E-10	1.97E-08	0.0396	0.0392	0.4005	0.0856	0.0431	0.0306
RRAGD	chr6	90121825	90121827	cg04879577	4.2246	10.0364	2.00E-09	5.27E-08	0.0416	0.0396	0.3958	0.0210	0.0250	0.0308
TBX2	chr17	59477018	59477020	cg22163056	5.1185	11.3161	2.39E-10	8.86E-09	0.0204	0.0197	0.3916	0.0466	0.0227	0.0117
SLC47A1	chr17	19437790	19437792	cg17332016	4.6368	10.1756	1.58E-09	4.29E-08	0.0302	0.0279	0.3890	0.0348	0.0339	0.0341
TEPP	chr16	58018879	58018881	cg07844931	5.5384	11.4297	2.00E-10	7.62E-09	0.0159	0.0143	0.3885	0.0133	0.0201	0.0114
DCBLD2	chr3	98620634	98620636	cg07906351	5.9518	7.5830	2.05E-07	2.79E-06	0.0105	0.0105	0.3858	0.0100	0.0131	0.0100
SCUBE2	chr11	9113166	9113168	cg10782380	4.7702	8.2696	5.18E-08	8.51E-07	0.0261	0.0241	0.3783	0.0280	0.0312	0.0113
SMOC1	chr14	70346220	70346222	cg15996043	3.3892	6.6998	1.31E-06	1.39E-05	0.0752	0.0722	0.3771	0.0551	0.0580	0.0263
ENO4	chr10	118609144	118609146	cg02909379	3.6950	12.3620	4.78E-11	2.38E-09	0.0566	0.0550	0.3750	0.0532	0.0638	0.0646
ZNF365	chr10	64133882	64133884	cg09906751	4.3542	16.3624	2.37E-13	3.99E-11	0.0335	0.0324	0.3740	0.0896	0.0400	0.0473
DNER	chr2	230578916	230578918	cg16365799	4.8184	9.2968	7.46E-09	1.61E-07	0.0248	0.0227	0.3727	0.0117	0.0265	0.0144
SCARF2	chr22	20792142	20792144	cg03143742	4.6360	12.4185	4.39E-11	2.22E-09	0.0266	0.0258	0.3710	0.0183	0.0374	0.0137
SH3RF3	chr2	109745515	109745517	cg22725460	4.0024	8.5825	2.83E-08	5.06E-07	0.0427	0.0420	0.3705	0.0383	0.0372	0.0412
PPAP2B	chr1	57110721	57110723	cg09763175	4.5006	8.1441	6.63E-08	1.05E-06	0.0316	0.0284	0.3697	0.0475	0.0369	0.0129
GRIN2A	chr16	10276579	10276581	cg09239744	4.4023	7.2563	4.02E-07	5.01E-06	0.0307	0.0295	0.3619	0.0789	0.0320	0.0187
RLTPR	chr16	67678825	67678827	cg03314195	3.6049	7.8308	1.24E-07	1.81E-06	0.0561	0.0543	0.3571	0.0318	0.0486	0.0187
CKB	chr14	103989335	103989337	cg13714067	4.4320	12.8808	2.23E-11	1.29E-09	0.0290	0.0279	0.3550	0.0574	0.0302	0.0252
CHRM1	chr11	62691250	62691252	cg04557452	5.4764	9.6967	3.64E-09	8.73E-08	0.0133	0.0127	0.3520	0.0478	0.0105	0.0100
N/A	chr17	35014411	35014413	cg08967106	5.1016	12.3247	5.05E-11	2.49E-09	0.0173	0.0167	0.3512	0.0165	0.0247	0.0100
BOC	chr3	112930781	112930783	cg05953927	5.7288	6.8758	8.99E-07	1.01E-05	0.0107	0.0106	0.3512	0.0214	0.0100	0.0100
N/A	chr1	76081271	76081273	cg03662422	4.0379	11.1032	3.37E-10	1.18E-08	0.0379	0.0371	0.3507	0.0432	0.0507	0.0537
N/A	chr16	22173764	22173766	cg01869826	5.4560	10.7594	5.90E-10	1.88E-08	0.0131	0.0126	0.3465	0.0162	0.0190	0.0100
N/A	chr13	100547248	100547250	cg14684709	4.5799	11.0244	3.82E-10	1.31E-08	0.0253	0.0240	0.3462	0.0279	0.0372	0.0150
AMZ1	chr7	2728343	2728345	cg17081778	5.6184	10.2632	1.36E-09	3.79E-08	0.0110	0.0109	0.3395	0.0156	0.0132	0.0115
ADAM22	chr7	87563907	87563909	cg05351302	3.9549	7.9076	1.06E-07	1.58E-06	0.0383	0.0375	0.3390	0.0382	0.0375	0.0493
N/A	chr17	35014437	35014439	cg00371418	3.5126	9.2527	8.09E-09	1.73E-07	0.0567	0.0534	0.3383	0.0610	0.0654	0.0555
CKB	chr14	103989314	103989316	cg14294629	5.5605	14.8260	1.58E-12	1.64E-10	0.0115	0.0112	0.3376	0.0368	0.0149	0.0101
IGSF22	chr11	18727773	18727775	cg12670477	3.8993	8.7403	2.10E-08	3.91E-07	0.0456	0.0388	0.3370	0.0473	0.0439	0.0274
RTN1	chr14	60337153	60337155	cg07544748	3.9460	8.4161	3.90E-08	6.65E-07	0.0385	0.0369	0.3345	0.0307	0.0352	0.0155
PPAP2B	chr1	57111498	57111500	cg05494069	3.7885	9.4733	5.42E-09	1.23E-07	0.0455	0.0416	0.3334	0.0832	0.0569	0.0481

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{norm}	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
RRAGD	chr6	90122496	90122498	cg19187185	3.3849	9.3540	6.72E-09	1.48E-07	0.0586	0.0561	0.3269	0.0512	0.0781	0.0573
GRASP	chr12	52400649	52400651	cg09889291	5.2518	9.8050	3.00E-09	7.43E-08	0.0144	0.0133	0.3266	0.0100	0.0111	0.0116
TUBA3FP pseudogene	chr22	21368602	21368604	cg06912515	5.6109	9.3912	6.29E-09	1.39E-07	0.0102	0.0102	0.3245	0.0144	0.0100	0.0100
TSPAN9	chr12	3309494	3309496	cg07987843	4.8133	12.7495	2.70E-11	1.51E-09	0.0187	0.0179	0.3211	0.0114	0.0158	0.0132
MDGA1	chr6	37663981	37663983	cg08494905	3.5289	8.1665	6.35E-08	1.01E-06	0.0490	0.0482	0.3208	0.0492	0.0608	0.0526
FEV	chr2	219849188	219849190	cg01157126	3.9585	10.9820	4.10E-10	1.39E-08	0.0355	0.0339	0.3193	0.0927	0.0453	0.0130
GRASP	chr12	52400529	52400531	cg08102256	5.0152	13.1377	1.55E-11	9.70E-10	0.0153	0.0149	0.3130	0.0191	0.0267	0.0128
SULF1	chr8	70405203	70405205	cg20169702	3.6043	9.7537	3.29E-09	8.02E-08	0.0439	0.0430	0.3101	0.0500	0.0601	0.0322
CTSF	chr11	66336124	66336126	cg17381565	5.5295	11.1987	2.89E-10	1.04E-08	0.0100	0.0100	0.3081	0.0144	0.0118	0.0100
N/A	chr3	46940346	46940348	cg21792155	3.0056	5.9016	7.68E-06	6.39E-05	0.0732	0.0693	0.3049	0.0780	0.0681	0.0504
CKB	chr14	103989207	103989209	cg02248320	4.5138	10.8238	5.31E-10	1.72E-08	0.0226	0.0207	0.3046	0.0303	0.0238	0.0117
RHBDL3	chr17	30592656	30592658	cg04493025	5.0470	6.0487	5.51E-06	4.81E-05	0.0142	0.0138	0.3030	0.0114	0.0148	0.0100
QPCT	chr2	37572395	37572397	cg12040278	3.0675	6.0440	5.57E-06	4.85E-05	0.0656	0.0643	0.3013	0.0509	0.0658	0.0575
MGA	chr15	41913565	41913567	cg15924985	3.5711	11.7538	1.20E-10	5.04E-09	0.0427	0.0422	0.3013	0.0404	0.0426	0.0516
SLC16A14	chr2	230931792	230931794	cg02821501	4.3285	7.0212	6.59E-07	7.70E-06	0.0236	0.0233	0.3003	0.0145	0.0189	0.0261
STOX2	chr4	184826313	184826315	cg07089892	3.5119	8.4247	3.83E-08	6.56E-07	0.0454	0.0440	0.3002	0.0253	0.0371	0.0276
STK33	chr11	8615750	8615752	cg15247111	3.0251	5.4308	2.26E-05	0.000160683	0.0690	0.0661	0.2994	0.0236	0.0419	0.0319
TUBA3FP psuedogene	chr22	21368764	21368766	cg01038149	3.4432	10.2741	1.33E-09	3.73E-08	0.0463	0.0460	0.2980	0.0440	0.0459	0.0469
FAM132A	chr1	1182423	1182425	cg06443533	3.7782	9.5952	4.36E-09	1.02E-07	0.0370	0.0348	0.2963	0.0580	0.0571	0.0847
RTN1	chr14	60337616	60337618	cg04147372	3.5535	5.7194	1.16E-05	9.09E-05	0.0443	0.0416	0.2959	0.0209	0.0279	0.0268
ENO4	chr10	118608472	118608474	cg09763162	3.6571	8.6997	2.26E-08	4.18E-07	0.0422	0.0381	0.2950	0.0340	0.0355	0.0418
TRANK1	chr3	36986153	36986155	cg06163735	3.0070	6.0535	5.46E-06	4.76E-05	0.0664	0.0654	0.2945	0.0591	0.0624	0.0589
ENOX1	chr13	44361063	44361065	cg13319488	3.4003	8.4919	3.37E-08	5.88E-07	0.0478	0.0467	0.2942	0.0315	0.0449	0.0360
FEV	chr2	219849134	219849136	cg15138883	5.1554	8.2396	5.50E-08	8.95E-07	0.0124	0.0122	0.2930	0.0207	0.0100	0.0153
RP11-266L9.5 lincRNA	chr16	25078438	25078440	cg00611485	3.7876	7.4336	2.78E-07	3.64E-06	0.0341	0.0332	0.2887	0.0248	0.0295	0.0261
C17orf64	chr17	58498870	58498872	cg10179315	2.5745	5.5150	1.86E-05	0.000136029	0.0962	0.0927	0.2857	0.0561	0.0765	0.0794
ALPL	chr1	21835619	21835621	cg17015803	3.1905	8.6740	2.38E-08	4.36E-07	0.0535	0.0531	0.2856	0.0818	0.0587	0.0413
C1orf51	chr1	150254476	150254478	cg11207983	4.1445	9.3304	7.02E-09	1.53E-07	0.0253	0.0246	0.2842	0.0432	0.0300	0.0255

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{norm}	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
TTYH1	chr19	54926733	54926735	cg23695687	4.3693	11.2096	2.84E-10	1.02E-08	0.0231	0.0207	0.2831	0.0688	0.0336	0.0187
SLC16A14	chr2	230931921	230931923	cg00455164	5.2150	8.3194	4.70E-08	7.82E-07	0.0113	0.0111	0.2830	0.0114	0.0135	0.0230
EID3	chr12	104697192	104697194	cg01857475	2.6514	6.0529	5.46E-06	4.77E-05	0.0883	0.0841	0.2818	0.0342	0.0394	0.0622
ASB8	chr12	48576855	48576857	cg12396999	3.2355	13.3319	1.18E-11	7.82E-10	0.0508	0.0499	0.2809	0.0630	0.0876	0.0489
HCP5 sense_overlapping	chr6	31431069	31431071	cg12034086	4.1806	10.4123	1.05E-09	3.06E-08	0.0242	0.0235	0.2807	0.0269	0.0285	0.0186
GBGT1	chr9	136039282	136039284	cg24340029	2.6461	8.2933	4.95E-08	8.18E-07	0.0848	0.0838	0.2804	0.0712	0.0829	0.0955
RRAGD	chr6	90122151	90122153	cg10651618	3.5217	6.9866	7.10E-07	8.20E-06	0.0440	0.0390	0.2788	0.0166	0.0639	0.0118
SYT5	chr19	55685291	55685293	cg07721458	4.0892	10.4624	9.68E-10	2.85E-08	0.0256	0.0244	0.2744	0.0338	0.0495	0.0274
CDK14	chr7	90226449	90226451	cg02759151	3.6159	10.9176	4.55E-10	1.51E-08	0.0357	0.0350	0.2727	0.0400	0.0327	0.0345
VANGL2	chr1	160370362	160370364	cg18398058	3.3587	6.6893	1.34E-06	1.42E-05	0.0430	0.0422	0.2691	0.0402	0.0430	0.0393
ISM2	chr14	77965283	77965285	cg16049707	5.0457	10.5625	8.18E-10	2.47E-08	0.0120	0.0117	0.2686	0.0100	0.0119	0.0109
TUBA3FP pseudogene	chr22	21368706	21368708	cg21014483	4.6553	7.9397	9.95E-08	1.50E-06	0.0162	0.0154	0.2668	0.0202	0.0177	0.0191
PNPLA3	chr22	44319578	44319580	cg04692420	4.4558	9.8897	2.59E-09	6.54E-08	0.0190	0.0177	0.2655	0.0176	0.0160	0.0119
CACNA1G	chr17	48636899	48636901	cg24280645	5.1749	14.2419	3.39E-12	2.96E-10	0.0104	0.0103	0.2632	0.0100	0.0136	0.0100
RRAGD	chr6	90122153	90122155	cg09700990	2.7536	6.6853	1.35E-06	1.43E-05	0.0698	0.0683	0.2626	0.0398	0.0813	0.0374
N/A	chr4	24473412	24473414	cg14910061	3.2357	9.6013	4.31E-09	1.01E-07	0.0455	0.0449	0.2621	0.0556	0.0590	0.0561
STK33	chr11	8615574	8615576	cg02541031	3.5111	6.2467	3.54E-06	3.28E-05	0.0371	0.0359	0.2620	0.0213	0.0359	0.0188
STK33	chr11	8615693	8615695	cg00393798	3.4487	5.0336	5.69E-05	0.000353999	0.0379	0.0368	0.2575	0.0163	0.0217	0.0204
TUBA3FP pseudogene	chr22	21368673	21368675	cg23000550	4.6830	8.0497	7.99E-08	1.24E-06	0.0172	0.0143	0.2573	0.0142	0.0179	0.0100
CCDC85A	chr2	56410779	56410781	cg11572305	2.6956	7.9680	9.40E-08	1.43E-06	0.0713	0.0698	0.2573	0.0588	0.0790	0.0554
CACNA2D2	chr3	50541010	50541012	cg20220436	3.3138	7.0507	6.19E-07	7.28E-06	0.0419	0.0409	0.2568	0.0293	0.0581	0.0141
GALNT16	chr14	69727351	69727353	cg05157120	2.9376	7.3828	3.09E-07	3.99E-06	0.0568	0.0559	0.2561	0.0490	0.0610	0.0699
NKX1-2	chr10	126138878	126138880	cg25531679	2.9972	9.6593	3.89E-09	9.24E-08	0.0531	0.0522	0.2533	0.0363	0.0563	0.0464
TEPP	chr16	58018971	58018973	cg05869503	5.1431	9.8084	2.98E-09	7.39E-08	0.0100	0.0100	0.2530	0.0165	0.0100	0.0108
BOC	chr3	112931125	112931127	cg06189303	4.3543	7.0673	5.98E-07	7.06E-06	0.0182	0.0178	0.2527	0.0217	0.0175	0.0132
B4GALNT1	chr12	58026475	58026477	cg21818749	4.9934	10.6400	7.19E-10	2.22E-08	0.0111	0.0111	0.2517	0.0342	0.0107	0.0100
SCUBE2	chr11	9113369	9113371	cg02783889	4.0217	9.7544	3.28E-09	8.01E-08	0.0229	0.0222	0.2474	0.0300	0.0265	0.0151
ZNF599	chr19	35263966	35263968	cg00259785	3.4955	6.6676	1.41E-06	1.48E-05	0.0334	0.0331	0.2455	0.0360	0.0474	0.0324

Appendix 7: Hypermethylation biomarker probes for BE and EAC

gencode	chr	start	end	probe	logfc	tstat	pval	adj_pval	β_{norm}	β_{base}	$\Delta\beta$	β_{duo}	β_{stom}	β_{blood}
LRRC43	chr12	122667626	122667628	cg21176475	3.6331	6.1937	3.99E-06	3.64E-05	0.0304	0.0296	0.2451	0.0275	0.0241	0.0308
SMAD9	chr13	37494041	37494043	cg03754180	2.8654	6.0744	5.21E-06	4.57E-05	0.0566	0.0548	0.2424	0.0540	0.0561	0.0437
RPL7P44 pseudogene	chr13	77554506	77554508	cg25163979	3.4808	8.4202	3.87E-08	6.61E-07	0.0335	0.0326	0.2408	0.0535	0.0514	0.0260
PLXND1	chr3	129324500	129324502	cg01060026	4.8684	8.6660	2.41E-08	4.41E-07	0.0115	0.0113	0.2388	0.0208	0.0100	0.0111
MAST4	chr5	65893335	65893337	cg26172694	5.0263	8.8942	1.57E-08	3.06E-07	0.0100	0.0100	0.2377	0.0109	0.0103	0.0100
PCLO	chr7	82792410	82792412	cg06129054	3.9476	9.2851	7.62E-09	1.64E-07	0.0224	0.0217	0.2336	0.0373	0.0285	0.0252
TOX2	chr20	42543406	42543408	cg23872756	4.3602	8.6709	2.39E-08	4.38E-07	0.0169	0.0159	0.2336	0.0144	0.0240	0.0113
AKT3	chr1	244014034	244014036	cg16546309	3.0670	6.1218	4.68E-06	4.17E-05	0.0444	0.0435	0.2325	0.0210	0.0403	0.0192
ABR	chr17	1082704	1082706	cg07464092	3.6904	5.8202	9.24E-06	7.48E-05	0.0278	0.0262	0.2316	0.0170	0.0249	0.0254
MYH10	chr17	8534125	8534127	cg01044722	4.8301	7.0756	5.88E-07	6.96E-06	0.0111	0.0110	0.2295	0.0100	0.0114	0.0107
CR2	chr1	207627580	207627582	cg09936645	4.6799	7.0231	6.57E-07	7.67E-06	0.0140	0.0122	0.2285	0.0100	0.0104	0.0100
ENO4	chr10	118608840	118608842	cg03871675	2.7984	11.7732	1.17E-10	4.92E-09	0.0534	0.0529	0.2268	0.0426	0.0462	0.0555
N/A	chr7	127807620	127807622	cg16511994	3.2933	7.0918	5.68E-07	6.76E-06	0.0370	0.0345	0.2248	0.0186	0.0316	0.0465
LRRC43	chr12	122667955	122667957	cg26953749	3.0535	6.9742	7.29E-07	8.39E-06	0.0426	0.0419	0.2243	0.0403	0.0345	0.0442
MKX	chr10	28031329	28031331	cg04923620	2.3721	7.0465	6.25E-07	7.34E-06	0.0779	0.0765	0.2237	0.0610	0.0676	0.0728
C9orf171	chr9	135285408	135285410	cg08987373	2.4646	7.2610	3.98E-07	4.97E-06	0.0710	0.0700	0.2235	0.0751	0.0696	0.0779
DRD2	chr11	113346206	113346208	cg20007462	2.4898	12.3338	4.98E-11	2.46E-09	0.0682	0.0675	0.2215	0.0554	0.0716	0.0561
MKX	chr10	28031090	28031092	cg01424281	3.1190	6.0132	5.97E-06	5.15E-05	0.0412	0.0389	0.2211	0.0222	0.0195	0.0279
NTRK1	chr1	156830079	156830081	cg03055346	3.3025	8.3070	4.82E-08	7.99E-07	0.0333	0.0328	0.2180	0.0261	0.0334	0.0213
STK33	chr11	8615578	8615580	cg12675870	2.5914	5.0911	4.97E-05	0.000315326	0.0603	0.0588	0.2148	0.0294	0.0490	0.0455
LMO2	chr11	33891204	33891206	cg03476291	2.2343	5.6539	1.35E-05	0.000103395	0.0825	0.0818	0.2136	0.0712	0.0786	0.0734
PDE2A	chr11	72353586	72353588	cg23186104	2.4587	6.0314	5.73E-06	4.97E-05	0.0678	0.0655	0.2125	0.0618	0.0802	0.0329
STK33	chr11	8615505	8615507	cg08788717	4.8219	6.6270	1.54E-06	1.60E-05	0.0100	0.0100	0.2122	0.0100	0.0100	0.0100
LRRC43	chr12	122667648	122667650	cg26211360	3.0302	7.0123	6.72E-07	7.82E-06	0.0416	0.0389	0.2094	0.0340	0.0443	0.0511
STEAP1	chr7	89783611	89783613	cg05797929	4.2563	8.8284	1.78E-08	3.39E-07	0.0151	0.0147	0.2070	0.0134	0.0124	0.0118
KLHL14	chr18	30351341	30351343	cg03513246	2.3861	5.5843	1.58E-05	0.000118681	0.0690	0.0673	0.2066	0.0567	0.0583	0.0689

Appendix 8: Pan-cancer targets for mutational profiling

Pan-cancer panel for mutational profiling. Samples were profiled across two separate runs, the first profiling 300 targets, the second including an additional 13 targets (total n=313, v2) as the panel was developed. ENST: Ensembl spliced transcript.

Gene	Ensembl Reference	Gene	Ensembl Reference
Cancer genes (full length) from other sources including Foundation Medicine, Illumina's TruSight Tumor Panel and TruSeq Amplicon			
LRP2	ENST00000263816	CCND1	ENST00000227507
CDKN2A (P16)	ENST00000579755	CCND2	ENST00000261254
RPA1	ENST00000254719	CCND3	ENST00000372991
BRCA1	ENST00000357654	CCNE1	ENST00000262643
BRCA2	ENST00000544455	CD79A	ENST00000221972
ATM	ENST00000278616	CD79B	ENST00000392795
PALB2	ENST00000261584	CDC73	ENST00000367435
DAXX	ENST00000374542	CDK12	ENST00000447079
ATRX	ENST00000373344	CDK4	ENST00000257904
NF1	ENST00000358273	CDK6	ENST00000424848
Ikaros/IKZF1	ENST00000331340	CDK8	ENST00000381527
WIF1	ENST00000286574	CDKN2B	ENST00000276925
SNIP1	ENST00000296215	CDKN2C	ENST00000371761
SMAD4	ENST00000342988	CEBPA	ENST00000498907
p53	ENST00000269305	CHEK1	ENST00000534070
MGMT	ENST00000306010	CHEK2	ENST00000382580
PTTG	ENST00000393964	CIC	ENST00000575354
EGF	ENST00000265171	CRKL	ENST00000411769
EGFR	ENST00000275493	CRLF2	ENST00000381567
BRAF	ENST00000288602	CTCF	ENST00000264010
HIF1 alpha	ENST00000539097	CTNNA1	ENST00000302763
bFGF/FGF2	ENST00000264498	DNMT3A	ENST00000264709
Ptd-FGFR4	ENST00000292408	DOT1L	ENST00000398665
E-cadherin/CDH1	ENST00000261769	EMSY/C11orf30	ENST00000334736
pNCAM	NCAM1	EMSY/C11orf30	ENST00000263253
MMP2	ENST00000219070	ERBB3	ENST00000267101
MMP9	ENST00000372330	ERG	ENST00000417133
NM23/NME2	ENST00000393193	ESR1	ENST00000440973
Hst/FGF4	ENST00000168712	EZH2	ENST00000320356
Topo II alpha/TOP2A	ENST00000423485	FAM123B (WTX)/AMER1	ENST00000330258
Galectin 3/LGALS3	ENST00000254301	FAM46C	ENST00000369448
VEGF	ENST00000372067	FANCA	ENST00000389301
COX2/PTGS2	ENST00000367468	FANCC	ENST00000289081
AKT1	ENST00000554581	FANCD2	ENST00000287647
PTEN	ENST00000371953	FANCE	ENST00000229769
HER-2/ERBB2	ENST00000406381	FANCF	ENST00000327470
KRAS	ENST00000256078	FANCG	ENST00000378643
NRAS	ENST00000369535	FANCL	ENST00000402135
BCL2	ENST00000398117	FGF10	ENST00000264664
adamts6	ENST00000381055	FGF14	ENST00000376131
crmp1	ENST00000324989	FGF19	ENST00000294312
dcamkl3	ENST00000416516	FGF23	ENST00000237837
ask	ENST00000265728	FGF3	ENST00000334134
CCNB1	ENST00000256442	FGF6	ENST00000228837

Appendix 8: Pan-cancer targets for mutational profiling

Gene	Ensembl Reference	Gene	Ensembl Reference
AURKB	ENST00000316199	FLT1	ENST00000282397
CENPE	ENST00000265148	FLT4	ENST00000261937
HMGA1	ENST00000447654	GID4 (C17orf39)	ENST00000268719
HMGA2	ENST00000403681	GNA13	ENST00000439174
WNT4	ENST00000290167	GPR124	ENST00000412232
WNT9a	ENST00000272164	GRIN2A	ENST00000396573
ALK	ENST00000389048	GSK3B	ENST00000316626
APC	ENST00000457016	HGF	ENST00000222390
FBXW7	ENST00000281708	IDH2	ENST00000330062
FGFR2	ENST00000358487	IGF1R	ENST00000268035
FOXL2	ENST00000330315	IKBKE	ENST00000367120
GNAQ	ENST00000286548	IL7R	ENST00000303115
GNAS	ENST00000371100	INHBA	ENST00000242208
KIT	ENST00000288135	IRF4	ENST00000380956
MAPK2	ENST00000215832	IRS2	ENST00000375856
MET	ENST00000318493	JAK1	ENST00000342505
MSH6	ENST00000234420	KAT6A	ENST00000396930
PDGFRA	ENST00000257290	KDM5A	ENST00000399788
PIK3CA	ENST00000263967	KDM5C	ENST00000375401
SRC	ENST00000373578	KEAP1	ENST00000171111
STK11	ENST00000326873	KLHL6	ENST00000341319
ABL1	ENST00000372348	LRP1B	ENST00000389484
MLH1	ENST00000231790	MAP2K1	ENST00000307102
RET (MEN2)	ENST00000355710	MAP2K2	ENST00000262948
ERBB4	ENST00000342788	MAP2K4	ENST00000353533
HNF1A	ENST00000257555	MAP3K1	ENST00000399503
MPL	ENST00000372470	MCL1	ENST00000369026
NOTCH1	ENST00000277541	MED12	ENST00000374080
SMARCB1	ENST00000263121	MEF2B	ENST00000424583
IDH1	ENST00000415913	MITF	ENST00000352241
NPM1	ENST00000296930	MLL2	ENST00000301067
SMO	ENST00000249373	MRE11A	ENST00000323929
JAK2	ENST00000381652	MSH2	ENST00000233146
JAK3	ENST00000458235	MUTYH	ENST00000372098
KDR	ENST00000263923	MYC	ENST00000377970
FLT3	ENST00000241453	MYCL1	ENST00000397332
VHL	ENST00000256474	MYCN	ENST00000281043
CSF1R	ENST00000286301	MYD88	ENST00000417037
GNA11	ENST00000078429	NF2	ENST00000338641
PTPN11	ENST00000351677	NFE2L2	ENST00000397062
RB1	ENST00000267163	NFKBIA	ENST00000216797
SMARCA4	ENST00000429416	NKX2-1	ENST00000354822
FOXA1	ENST00000250448	NOTCH2	NOTCH2
ASXL1	ENST00000375687	NTRK1	ENST00000524377
CYP2D6	ENST00000360608	NTRK2	ENST00000376214
DDR2	ENST00000367922	NTRK3	ENST00000360948
KDM6A	ENST00000377967	NUP93	ENST00000308159
NKX3-1	ENST00000380871	PAK3	ENST00000360648
ZFHX3	ENST00000268489	PBRM1	ENST00000394830
JUN	ENST00000371222	PDGFRB	ENST00000261799
MTOR	ENST00000361445	PDK1	ENST00000410055
RAF1	ENST00000251849	PIK3CG	ENST00000359195
GATA1	ENST00000376670	PIK3R1	ENST00000521381
EPHA3	ENST00000336596	PIK3R2	PIK3R2
CREBBP	ENST00000262367	PPP2R1A	ENST00000322088

Gene	Ensembl Reference	Gene	Ensembl Reference
ARID1A	ENST00000324856	PRDM1	ENST00000369096
SPOP	ENST00000393331	PRKDC	ENST00000314191
AR	ENST00000374690	PTCH1	ENST00000331920
MDM2	ENST00000462284	RAD50	ENST00000265335
UGT1A1	ENST00000305208	RAD51	ENST00000382643
EPHB1	ENST00000398015	RARA	ENST00000254066
MLL1	ENST00000534358	RICTOR	ENST00000357387
FGFR3	ENST00000340107	RNF43	ENST00000584437
CHD1	ENST00000284049	RPTOR	ENST00000306801
MDM4	ENST00000367182	RUNX1	ENST00000300305
EPHA5	ENST00000273854	SETD2	ENST00000409792
MLL3	ENST00000262189	SF3B1	ENST00000335508
GATA2	ENST00000341105	SMAD2	ENST00000262160
GATA3	ENST00000379328	SOCS1	ENST00000332029
HRAS	ENST00000451590	SOX10	ENST00000396884
AKT2	ENST00000424901	SPEN	ENST00000375759
AKT3	ENST00000263826	STAG2	ENST00000218089
ARAF	ENST00000377045	STAT4	ENST00000392320
ARFRP1	ENST00000359715	SUFU	ENST00000369902
ARID2	ENST00000334344	TET2	ENST00000540549
ATR	ENST00000350721	TGFBR2	ENST00000359013
AURKA	ENST00000395909	TNFAIP3	ENST00000237289
AXL	ENST00000301178	TNFRSF14	ENST00000355716
BAP1	ENST00000460680	TOP1	ENST00000361337
BARD1	ENST00000260947	TSC1	ENST00000298552
BCL2L2	ENST00000250405	TSC2	ENST00000219476
BCL6	ENST00000406870	TSHR	ENST00000541158
BCOR	ENST00000342274	WISP3	WISP3
BCORL1	ENST00000540052	WT1	ENST00000332351
BLM	ENST00000355112	XPO1	ENST00000406957
BRIP1	ENST00000259008	ZNF217	ENST00000371471
BTK	ENST00000308731	ZNF703	ENST00000331569
CARD11	ENST00000396946	FBX4	ENST00000281623
CBFB	ENST00000412916	SOS1	ENST00000402219
CBL	ENST00000264033		
Familial pituitary genes			
AIP	ENST00000279146	SDHA	ENST00000264932
MEN1	ENST00000337652	SDHB	ENST00000375499
p27 (KIP1)/CDKN1B	ENST00000228872	SDHC	ENST00000367975
PRKAR1A (Carney Complex)	ENST00000536854	SDHD	ENST00000375549
Pituitary development (embryonic) genes			
PIT1	ENST00000350375	PAX7	ENST00000420770
LHX2	ENST00000373615	FGF8	ENST00000320185
LHX3	ENST00000371746	FGFR1	ENST00000447712
LHX4	ENST00000263726	BMP4	ENST00000245451
SOX2	ENST00000325404	BMP8	ENST00000372827
SOX3	ENST00000370536	SHH	ENST00000297261
SOX9	ENST00000245479	GLI2	ENST00000452319
β-catenin/CTNNB1	ENST00000349496	TCF7L1	ENST00000282111
HESX1	ENST00000295934	SIX6	ENST00000327720
OTX2	ENST00000339475	SIX3	ENST00000260653
PROP1	ENST00000308304	KAL1	ENST00000262648
PAX5	ENST00000358127	CHD7	ENST00000423902
PAX6	ENST00000419022		

Appendix 8: Pan-cancer targets for mutational profiling

Gene	Ensembl Reference	Gene	Ensembl Reference
Additional genes to screen in Pan-cancer panel v2			
MCC	ENST00000408903	RYR1	ENST00000359596
JRK	ENST00000507178	CACNA1S	ENST00000362061
CCAR2	ENST00000308511	CDKN2A (P14ARF isoform)	ENST00000361570
SIRT1	ENST00000212015	DDX3X	ENST00000399959
BCL9	ENST00000234739	RAC1	ENST00000356142
PMS2	ENST00000265849	TERT incl. promoter region	ENST00000310581
EPCAM	ENST00000263735		

REPRINTS OF PUBLICATIONS ARISING DURING MY THESIS CANDIDATURE

Levert-Mignon A, Bourke MJ, Lord SJ, Taylor AC, Wettstein AR, Edwards M, Botelho NK, Sonson R, Jayasekera C, Fisher OF, **Thomas ML**, Macrae F, Hussey DJ, Watson D, Lord RV. Changes in gene expression of neo-squamous mucosa after endoscopic treatment for dysplastic Barrett's esophagus and intramucosal adenocarcinoma. United European Gastroenterol. J. 2016, May 19. Doi:10.1177/2050640616650794.

Fisher OM, Levert-Mignon AJ, Lord SJ, Botelho NK, Freeman AK, **Thomas ML**, Falkenback D, Wettstein A, Whiteman DC, Bobryshev YV, Lord RV. High Expression of Cathepsin E in Tissues but not Blood of Patients with Barrett's Esophagus and Adenocarcinoma. Ann Surg Oncol. 2015, July; 22(7):2431-8. doi: 10.1245/s10434-014-4155-y.

Whiteman DC, Appleyard M, Bahin FF, Bobryshev YV, Bourke MJ, Brown I, Chung A, Clouston A, Dickins E, Emery J, Eslick GD, Gordon LG, Grimpen F, Hebbard G, Holliday L, Hourigan L, Kendall BJ, Lee EY, Levert A, Lord RV, Lord SJ, Maule D, Moss A, Norton I, Olver I, Pavey D, Raftopoulos S, Rajendra S, Schoeman M, Singh R, Sitas F, Smithers BM, Taylor A, **Thomas ML**, Thomson I, To H, von Dincklage J, Vuletich C, Watson DI and Yusoff IF. Australian clinical practice guidelines for the diagnosis and management of Barrett's Esophagus and Early Esophageal Adenocarcinoma. J. Gastroenterol. Hepatol. 2015, May; 30(5):804-20. doi: 10.1111/jgh.12913.

Fisher OM, Levert-Mignon AJ, Lord SJ, Lee-Ng KKM, Botelho NK, Falkenback D, **Thomas ML**, Bobryshev YV, Whiteman DC, Brown DA, Breit SN, Lord RV. MIC-1/GDF15 in Barrett's Esophagus and Esophageal Adenocarcinoma. Br. J. Cancer. 2015, Apr 14; 112(8)1384-91. doi: 10.1038/bjc.2015.100.

UNITED EUROPEAN GASTROENTEROLOGY ueg journal

Original Article

United European Gastroenterology Journal
0(0) 1–8

DOI: 10.1177/2050640616650794
ueg.sagepub.com
SAGE

Changes in gene expression of neo-squamous mucosa after endoscopic treatment for dysplastic Barrett's esophagus and intramucosal adenocarcinoma

Angelique Levert-Mignon[1], Michael J Bourke[2], Sarah J Lord[1,3], Andrew C Taylor[4], Antony R Wettstein[5], Melanie Edwards[3,6], Natalia K Botelho[1], Rebecca Sonson[2], Chatura Jayasekera[4], Oliver M Fisher[1], Melissa L Thomas[1,3], Finlay Macrae[7,8], Damian J Hussey[9,10], David I Watson[9,10] and Reginald V Lord[1,3,5]

Abstract

Background: Endoscopic therapy, including by radiofrequency ablation (RFA) or endoscopic mucosal resection (EMR), is first line treatment for Barrett's esophagus (BE) with high-grade dysplasia (HGD) or intramucosal cancer (IMC) and may be appropriate for some patients with low-grade dysplasia (LGD).

Objective: The purpose of this study was to investigate the molecular effects of endotherapy.

Methods: mRNA expression of 16 genes significantly associated with different BE stages was measured in paired pre-treatment BE tissues and post-treatment neo-squamous biopsies from 36 patients treated by RFA (19 patients, 3 IMC, 4 HGD, 12 LGD) or EMR (17 patients, 4 IMC, 13 HGD). EMR was performed prior to RFA in eight patients. Normal squamous esophageal tissues were from 20 control individuals.

Results: Endoscopic therapy resulted in significant change towards the normal squamous expression profile for all genes. The neo-squamous expression profile was significantly different to the normal control profile for 11 of 16 genes.

Conclusion: Endotherapy results in marked changes in mRNA expression, with replacement of the disordered BE dysplasia or IMC profile with a more ''normal'' profile. The neo-squamous mucosa was significantly different to the normal control squamous mucosa for most genes. The significance of this finding is uncertain but it may support continued endoscopic surveillance after successful endotherapy.

Keywords

Barrett's esophagus, high-grade dysplasia, esophageal adenocarcinoma, esophageal neoplasms, radiofrequency ablation, endoscopic mucosal resection

Received: 8 December 2015; accepted: 27 April 2016

[1]Gastroesophageal Cancer Research Program, St Vincent's Centre for Applied Medical Research, Sydney, NSW, Australia
[2]Department of Gastroenterology and Hepatology, Westmead Hospital, Westmead, NSW, Australia
[3]School of Medicine, University of Notre Dame, Sydney, NSW, Australia
[4]Department of Gastroenterology, St Vincent's Hospital, Melbourne, VIC, Australia
[5]Diagnostic Endoscopy Centre, St Vincent's Clinic, Sydney, NSW, Australia
[6]Department of Histopathology, Douglass Hanly Moir Pathology, Sydney, NSW Australia
[7]Department of Colorectal Medicine and Genetics, The Royal Melbourne Hospital, Parkville, VIC, Australia
[8]Department of Medicine, The University of Melbourne, Parkville, VIC, Australia
[9]Department of Surgery, Flinders University, Flinders Medical Centre Bedford Park, SA, Australia
[10]Flinders Centre for Cancer Prevention and Control, Flinders University, Bedford Park, SA, Australia

Corresponding author:
Reginald V. Lord, Suite 606, St Vincent's Clinic, 438 Victoria St, Darlinghurst, NSW 2010, Australia.
Email: rvlord@stvincents.com.au

Introduction

Barrett's esophagus (BE) is a premalignant condition in which the normal squamous lining of the lower esophagus is replaced by an intestinal metaplastic (IM) columnar epithelium in response to prolonged severe gastro-esophageal reflux. BE is the major risk factor for esophageal adenocarcinoma (EAC), a cancer with a high case fatality ratio and a rapidly rising incidence.[1] The progression from normal esophagus to BE and adenocarcinoma is thought to involve a complex, multistep process, from IM to low-grade dysplasia (LGD), high-grade dysplasia (HGD), early intramucosal cancer (IMC), to invasive EAC.

Intervention is recommended for patients with HGD or IMC, based on the estimated 7–19% yearly risk of EAC developing in patients with HGD.[1] Endoscopic therapy has replaced esophagectomy as the preferred first-line treatment for most patients with HGD/IMC, as it avoids the morbidity and mortality associated with esophagectomy, preserves the esophagus, and has equivalent survival outcomes.[2,3] Guidelines[1,4,5] have recommended endoscopic mucosal resection (EMR) for visible lesions and radiofrequency ablation (RFA) of flat mucosae, including after EMR. Complete Barrett's excision (CBE) endoscopic resection is an alternative for shorter Barrett's segments.[2] More recently, endoscopic therapy has been recommended for patients with persistent and confirmed (by two expert gastrointestinal pathologists, in two or more endoscopies) multifocal LGD.[6,7]

Eradication of cancer, dysplasia, and all BE is reported in up to 94% of patients treated with RFA or EMR,[3,8–10] The risk of EAC development is also significantly reduced.[11] The rate of progression to EAC was one per 181 patient-years (0.55%/patient-years) in a multicenter US study, at three years after RFA or RFA and EMR combined treatment for dysplastic BE.[8] In another study, the cancer-related mortality rate was 0.2% in EMR treated patients with IMC after five years.[3]

Although RFA and EMR have proven to be safe and effective in at least the medium term,[9,12,13] there are reports of recurrence.[13] The durability over decades, which is relevant for this disease, and underlying molecular effect, remains unknown. It has previously been shown that the altered mRNA expression of certain genes is associated with different stages of the Barrett's to adenocarcinoma sequence.[14] By comparing gene expression in the tissue biopsies of dysplastic Barrett's or IMC mucosa before endoscopic therapy and in the normal-appearing neo-squamous mucosa post-treatment, we evaluated the molecular effect of endoscopic treatment.

Methods

Patients

Patients in the treatment group undergoing RFA, EMR, or combination of RFA plus EMR for the treatment of histopathologically confirmed BE with dysplasia or IMC were invited to participate in this prospective multi-center study. The treatment selected was at the discretion of the endoscopist. BE length was recorded using the Prague classification. Post-treatment biopsies were taken from the macroscopically normal appearing neo-squamous mucosa from the same area as the pre-treatment BE, as measured by distance from the incisors.

A control group consisted of individuals with the typical reflux symptoms of heartburn or regurgitation but without a history of current or past macroscopic reflux esophagitis (RE) or BE (non RE/BE). Inclusion criteria for both groups were age $\geq$18 years and ability to give informed consent. Approval for the study was obtained from the Human Research Ethics Committee at each center and participants provided written informed consent.

Endoscopic treatment

Radiofrequency ablation. After removal of visible lesions where present by EMR, the abnormal mucosa was ablated by RFA using either a circumferential balloon catheter or a flat plate device (BARRx Medical/Covidien, Inc., Sunnyvale California, USA). The radiofrequency energy was delivered to the Barrett's mucosa (12 J/cm^2, 40 W/cm^2) twice in sequence. The device was removed and cleaned between applications, and the ablated epithelium was cleaned by irrigation or scrapped off with the edge of the device.

Endoscopic mucosal resection. The irregular mucosa was resected using the Duette multiband mucosectomy system (Cook Medical, Bloomington, Indiana, USA). The mucosa is lifted by aspiration, ligated to form a pseudopolyp, and resected by electrocautery, as described previously.[10] Both RFA and EMR were performed in single or multiple sessions, depending on the extent of BE.

Tissue specimens

From review of the histopathological reports of routine hematoxylin and eosin-stained (H&E) tissue sections, archival formalin-fixed, paraffin-embedded (FFPE) esophageal tissue samples were obtained from the study centers. The worst histopathological grade of BE/IMC was selected for the pre-treatment dysplastic

BE or IMC study tissues. The post-treatment samples were matched neo-squamous mucosa collected from the same area as the previous BE mucosa.

RNA isolation

Two 7 µm unstained sections cut from each FFPE sample block were used for RNA extraction. At least 55 ng total RNA was isolated by a column-based purification method using the Ambion RecoverAll Total Nucleic Acid Isolation kit for FFPE, Cat # AM1975 (Life Technologies, Carlsbad, California, USA) or the QIAGEN RNeasy FFPE kit, Cat # 744404 (Qiagen, Valencia, California, USA), according to the manufacturer's protocol. RNA purity and concentration was measured using a NanoPhotometer (Implen, Westlake Village, California, USA).

mRNA quantification by multiplex tandem polymerase chain reaction (MT-PCR)

Sixteen genes significantly differentially expressed at the mRNA level in BE and EAC compared to normal squamous esophagus were selected from previous studies.[14,15] Full details of the MT-PCR methods were reported previously.[14] In brief, mRNA expression levels of the genes of interest and the internal reference gene, NONO (''non-POU domain containing, octamer-binding''; NM_007363), were measured in duplicate with pre- and post-treatment tissues assessed simultaneously. MT-PCR was performed using a real-time quantitative PCR system (Rotor-Gene RG6000, Qiagen, Valencia, California, USA). Primers for study genes and NONO were designed using Primer 3 software; the size of the ''inner'' amplicon was restricted to 70–90 bp and the ''outer'' amplicon to <150 bp. All primer pairs spanned an intron-exon boundary and the products were evaluated on a Bioanalyser DNA separation chip for the correct size (Agilent, Santa Clara, California, USA).

MT-PCR was performed in two steps. In the first step the RNA was converted into cDNA and amplified using multiplexed gene specific primers (''outer'' primers). In the second step the product from step one was used as a template for PCRs in a 72-well disc containing lyophilized single-gene primers (''inner'' primers) in each well. ''Outer'' primer mix was prepared by adding to one single tube 1 µl of each primer (forward and reverse) of all genes to 53 µl RNAse free diethylpyrocarbonate (DEPC) water to a total 125 µl, and they were lyophilized in 0.2 ml tubes. ''Inner'' primer mixes were prepared in different tubes (for each gene) by adding 4 µl of each primer (forward and reverse) into 424 µl of DEPC water.

Statistical analysis

The mRNA relative expression values were measured as the ratio of the absolute expression values of each target gene to the expression of the reference gene NONO, set to a fixed level (10,000). Gene expression values were not normally distributed, and therefore are summarized as medians with the 25th–75th interquartile range (IQR).

To identify genes differentially expressed post-treatment compared to pre-treatment values, unpaired (all subjects) and paired (subset of subjects with pre- and post-treatment samples) analyses were performed using the Wilcoxon rank-sum and signed rank test respectively. The Wilcoxon rank-sum test was also used to compare gene expression in normal squamous control versus post-treatment neo-squamous. Fold change was calculated to describe the magnitude of the change in gene expression levels pre- and post-treatment. All data analyses were performed using the SAS software (version 9.3).

Results

As shown in Table 1, 36 endotherapy patients (19 RFA, 17 EMR) and 20 control individuals were enrolled in the study. Most patients in both groups were male. In the RFA group, eight patients (42%) underwent focal EMR before RFA to remove nodular lesions containing IMC (three patients), HGD (four patients) or LGD (one patient). The remaining 11 patients were treated by RFA alone. In the EMR treatment group, five patients had CBE with complete eradication of BE in a single session; the remaining patients had stepwise EMR over more than one treatment session. Table 1 shows that EMR was performed for HGD or IMC, whereas the patients treated by RFA mostly (63%) had LGD in the untreated or post-EMR BE.

Maximal BE length was 6 cm pre-RFA and 9 cm pre-EMR, with a median of 2 cm for both treatment groups, and RFA or EMR was performed in up to three sessions at 2–3 month intervals. The median interval between the last treatment session and the post-treatment neo-squamous biopsy was three months for RFA and six months for EMR.

At the time of neo-squamous biopsy for this study, 17 of 19 (89%) RFA treated patients and 14 of 17 (82%) EMR patients had complete eradication of dysplasia; with complete eradication of IM in 6/19 (32%) RFA and 10/17 (59%) EMR patients. Subsequent to the post-treatment study biopsy, dysplasia was eradicated in all patients apart from one EMR patient who chose not to have further treatment because of advanced age, and IM has been eradicated in 58%

Table 1. Patients, pathology and treatment data

	RFA ($n=19$)	EMR ($n=17$)	No BE normal controls ($n=20$)
Sex, *n* (%)			
Males	17 (90%)	15 (88%)	17 (85%)
Females	2 (10%)	2 (12%)	3 (15%)
Age, years, median (range)	69 (52–84)	62 (31–82)	55 (33–77)
Length of BE (Prague Classification), cm, median (range)			
Circumferential (C)	0 (0–3)[a]	0 (0–5)	
Maximal (M)	2 (0.5–6)[a]	2 (0.5–9)	
Histological grade pre-treatment, *n* (%)			
Low-grade dysplasia	12 (63%)[a]	0 (0%)	
High-grade dysplasia	4 (21%)[a]	14 (76%)	
Intramucosal cancer	3 (16%)[a]	4 (24%)	
Median time, last treatment session to post-treatment biopsy (range), months	3 (1–15)	6 (1–57)	
No. of treatment sessions, median (range)	2 (1–3)	2 (1–3)	

BE: Barrett's esophagus; EMR: endoscopic mucosal resection; RFA: radiofrequency ablation.
[a]Pre-RFA, post-EMR wherever applicable.

Table 2. Difference in gene expression, pre- versus post-endoscopic treatment

	Pre-treatment BE			Post-treatment neo-squamous				
Gene	*n*	Median expression (IQR)		*n*	Median expression (IQR)		*p*-value[a]	Fold change
Downregulated								
TSPAN8	34	236.5	(8.9–2075.2)	25	3.8	(0.3–16.3)	<0.0001	−62
TSPAN1	36	13,219.4	(19.8–32,268.7)	33	226.6	(1.5–574.8)	0.003	−58
CTSE	36	15,185.0	(8383.0–23,798.0)	18	402.9	(66.9–5434.8)	0.0001	−38
MMP7	31	355.2	(127.2–556.7)	25	17.4	(4.5–73.7)	<0.0001	−20
MMP1	34	21.1	(9.2–92.3)	26	3.1	(1.1–5.7)	<0.0001	−7
COX-2	36	219.9	(92.7–559.3)	34	49.0	(19.8–159.2)	0.0005	−5
ODC1	36	4780.1	(3005.0–5798.1)	36	1434.9	(1044.2–2109.0)	<0.0001	−3
CD151	36	1657.6	(1241.8–2212.7)	33	574.8	(391.1–872.4)	<0.0001	−3
SPARC	36	15,177.1	(13,566.4–36,100.0)	36	5180.1	(4206.4–11,393.2)	<0.0001	−3
RARA	30	223.1	(131.3–283.7)	24	95.4	(69.1–154.5)	0.0004	−2
Upregulated								
ADH7	36	43.5	(11.3–172.6)	36	556.7	(294.4–902.8)	<0.0001	13
KRT4	36	32,854.0	(12,736.1–109,360.5)	36	251,647.6	(134,571.2–434,243.6)	<0.0001	8
RARG	33	108.3	(55.2–212.6)	33	674.9	(430.1–930.3)	<0.0001	6
SERPINB2	33	53.5	(6.2–98.4)	36	261.9	(110.2–1349.9)	<0.0001	5
PITX1	36	3252.2	(901.4–11,746.6)	36	10,505.4	(6280.0–80,569.9)	<0.0001	3
TP73L	33	8.1	(4.2–27.3)	35	23.0	(8.1–58.9)	0.02	3

IQR: interquartile range, [a]Wilcoxon test.

and 88% of RFA and EMR patients, respectively. There was no sub-squamous BE in any of the neo-squamous or normal squamous tissues.

Twenty individuals were enrolled in the control group. Normal esophagus biopsies were obtained from the distal esophagus in 11 individuals and from the proximal (~25 cm from incisors) esophagus in nine individuals.

Table 2 shows unpaired relative mRNA expression values for the 16 study genes in dysplastic Barrett's or

Table 3. Difference in gene expression: normal squamous versus neo-squamous

	Normal squamous			Neo-squamous			
Gene	*n*	Median expression (IQR)		*n*	Median expression (IQR)		*p*-value[a]
CD151	20	40.0	(12.4–141.3)	33	574.8	(391.1–872.4)	<0.0001
SPARC	20	1345.7	(259.7–3051.6)	36	5180.1	(4206.4–11393.2)	<0.0001
TP73L	20	246.6	(206.8–356.9)	35	23.0	(8.1–58.9)	<0.0001
PITX1	20	4458.0	(2770.1–8831.0)	36	10,505.4	(6280.0–80,569.9)	0.0002
ADH7	20	1370.1	(598.7–2572.7)	36	556.7	(294.4–902.8)	0.001
RARA	20	40.1	(14.3–90.9)	24	95.4	(69.1–154.5)	0.009
MMP7	20	100.3	(33.7–199.2)	25	17.4	(4.5–73.7)	0.0095
CTSE	12	37.8	(8.0–132.2)	18	402.9	(66.9–5434.8)	0.01
SERPINB2	20	1395.4	(996.0–2953.7)	36	261.9	(110.2–1349.9)	0.01
RARG	20	238.8	(153.5–598.5)	33	674.9	(430.6–930.3)	0.02
MMP1	17	8.6	(2.7–34.1)	26	3.1	(1.1–5.7)	0.02
TSPAN1	16	29.3	(5.2–60.1)	33	226.6	(1.5–574.8)	0.25
TSPAN8	10	8.8	(5.4–15.4)	25	3.8	(0.3–16.3)	0.32
ODC1	20	1370.1	(310.0–2147.3)	36	1434.9	(1044.2–2109.0)	0.33
KRT4	19	281,526	(9455–737,308)	36	251,647.6	(134,571.2–434,243.5)	0.41
COX-2	20	51.9	(21.1–87.0)	34	49.0	(19.8–159.2)	0.73

IQR: interquartile range, [a]Wilcoxon test.

IMC pre-endoscopic treatment tissues and matched neo-squamous tissue samples post-treatment in the 36 patients. Endoscopic therapy resulted in a highly significant change in median values for all 16 genes (Table 2). This was verified by paired analysis for patients with acceptable PCR results pre- and post endotherapy and the changes remained significant for all genes (data not shown). Ten genes were down-regulated and six genes were up-regulated in the normal neo-squamous mucosa after endoscopic treatment; all changes were from a Barrett's-associated profile towards a normal squamous epithelium profile.

Table 3 shows the relative gene expression levels in the neo-squamous epithelium compared to true normal squamous tissue from the 20 control individuals with gastro-esophageal reflux disease symptoms, but no RE/BE. The neo-squamous mucosa was significantly different to true normal squamous mucosa for 11/16 (69%) genes. This difference was most marked for CD151, SPARC and TP73L ($p < 0.0001$, shown graphically in Figure 1). Figure 1 also shows data for three genes with no significant difference between neo-squamous esophagus and true normal esophagus.

Comparing neo-squamous biopsies taken less than three months post-treatment versus more than three months post-treatment, there was no significant difference found except that SPARC mRNA expression was significantly lower in the greater than three months follow-up tissue cohort ($p = 0.0159$; data no shown)

Discussion

This study shows that there are marked changes in the relative mRNA expression levels of selected genes after RFA or EMR for the treatment of dysplastic BE or IMC. These changes are, as expected, towards a more "normal" squamous esophagus profile from non-BE patients. The mRNA expression in the neo-squamous mucosa post-treatment is not the same as found in the normal squamous mucosa, however, despite the neo-squamous mucosa being histopathologically indistinguishable from normal squamous epithelia.

Our findings indicate molecular as well as macro- and microscopic reversal of BE by endoscopic therapy. The expression of genes which have previously shown to be increased in a stepwise fashion from normal squamous esophagus to BE to EAC are significantly down-regulated by endotherapy, whereas those genes underexpressed in BE and EAC compared to normal mucosa are increased in expression after endotherapy. Although our study is the first to report mRNA expression changes, previous studies using different laboratory approaches have been reported.[16] Pouw et al. found no abnormal immunohistochemical (IHC) expression for Ki-67 and p53, and no numerical chromosomal abnormalities in the neo-squamous epithelia of 22 patients successfully treated with RFA for HGD or IMC.[17] Most of the patients (73%) in that study were treated with EMR before RFA for visible lesions, and salvage EMR was used on 18% of the

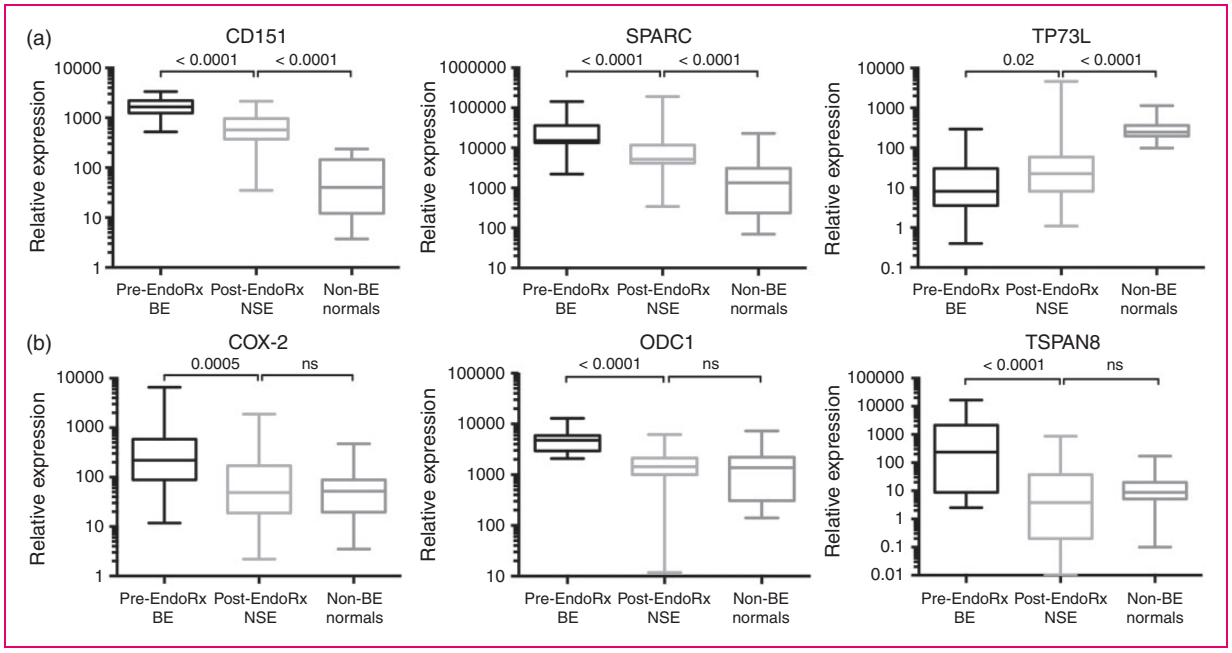


Figure 1. Endoscopic therapy resulted in significant changes in relative mRNA expression levels pre- and post-treatment for Barrett's esophagus (BE) with dysplasia or intramucosal cancer, compared to true normal squamous mucosa from control individuals with reflux symptoms but no BE. (a) Neo-squamous mucosa is significantly different to normal ($p < 0.0001$). (b) Neo-squamous mucosa difference is not significant (ns) compared to normal. Box plots show median (heavily longitudinal bar) and interquartile range (box).
EndoRx: Endotherapy; NSE: Neo-squamous epithelium.

patients after five RFA sessions to achieve complete eradication of BE. Krishnan et al., using IHC and Western blot methods, showed similar β-catenin expression in the neo-squamous and normal squamous mucosa at 12 months after successful RFA.[18] Other studies have shown persistent genetic abnormalities in remnant BE after photodynamic therapy (PDT) or argon plasma coagulation (APC).[19–21]

Comparing the gene expression profile of the neo-squamous mucosa with the normal squamous mucosa from individuals with typical reflux symptoms but no history of RE or BE, we found a significant difference for most genes. The relevance of this finding is unclear; we discuss three possible interpretations here. One interpretation is that this reflects ongoing wound healing, although only a minority of our selected genes (such as COX-2, matrix metallopeptidases 1 and 7) is clearly involved in wound healing. There was also no important change in the findings when we compared early post-treatment results (<3 months after endotherapy) with later post-treatment results (>3 months), suggesting that wound healing does not explain our results.

A second interpretation is that the differences in gene expression between the neosquamous and the normal squamous mucosa reflects a degree of molecular abnormality that is found even in the squamous mucosa in patients with BE. Brabender et al., for example, found a widespread carcinogenic field effect, measured in RNA quantification as in our study, in the normal squamous esophageal epithelia in patients with either BE or Barrett's adenocarcinoma.[22] In this respect the ideal design for our study would have included normal pre-treatment squamous esophagus tissues from the patients with BE, but we lacked the biopsy samples to do this.

A third interpretation is that patients with dysplastic BE/IMC retain some risk of disease persistence or recurrence, even after successful endoscopic therapy. This further suggests that ongoing surveillance after successful endotherapy is warranted, especially in younger patients. In keeping with this, Lewis et al. found raised cell proliferation (Ki-67) and COX-2 protein expression by IHC in buried subsquamous glands after APC. They interpreted this finding as making it unclear whether the risk of cancer is adequately reduced by ablation, with potential implications for patient follow up.[23] Similarly, Dijckmeester et al. found significantly higher expression of the microRNA-143 in neo-squamous after APC compared to normal squamous from control subjects, although expression of CK-8, CK-14, and microRNA-205 was similar.[24]

Clinical studies also suggest the need for ongoing surveillance and optimal reflux control after endoscopic therapy. Disease recurrence has been reported after complete eradication of Barrett's at variable rates. In a Netherlands cohort study, IM was present in 10% of patients treated with RFA after EMR for visible nodules

at five years after treatment.[25] Others report worse outcomes, including 33% BE recurrence rates at two-year follow-up after EMR and RFA,[12] 5% recurrence of IM per year after RFA,[26] and 14.5% recurrence of neoplasia (HGD or EAC) after approximately two years for EMR.[3] Cancer can recur even five or more years after successful endotherapy.[3] There are several clinical factors associated with worse response to endotherapy, including ongoing acid reflux exposure (which is usual in BE patients treated by PPIs), longer Barrett's segment, and a longer history of dysplastic BE.[27–29]

Genetic biomarkers could play a role in predicting response to endoscopic treatment.[16] A lower response to endoscopic treatment has been reported in patients with multiple chromosomal gains (gain of two or more locus-specific probes to MYC, p16, HER-2/neu and ZNF217, evaluated by fluorescence in situ hybridization (FISH)) in the dysplastic Barrett's epithelium.[30] After PDT, p16 allelic loss, also detected by FISH, was found to predict loss of dysplasia.[31]

Our study was prospective, used biopsies evaluated by H&E (rather than an adjacent biopsy with unknown pathology), and by simultaneously running pre- and post- treatment biopsies we limited the possibility of a ''batch effect''. Despite these methodological advantages, we acknowledge some limitations. Some neo-squamous biopsies were obtained pre-complete BE eradication. Consequently, it is possible (but unknown) if the remnant BE may effect the gene expression of the neo-squamous mucosa. Our normal squamous samples also include biopsies at various levels above the gastro-esophageal junction, which has been reported to influence gene expression,[32] although there was no significant difference in expression in distal compared to proximal esophagus biopsies in our study (data not shown). We did not compare the mRNA expression changes after EMR compared to RFA because of the small number of patients in each group and the difference in severity of Barrett's disease: most of the patients undergoing EMR had IMC or HGD whereas patients undergoing RFA had mostly LGD (after EMR treatment of IMC or HGD in some cases).

Our study found that the abnormal gene expression present at baseline in patients with dysplastic Barrett's or IMC is altered after endotherapy towards a normal esophagus expression profile. This alteration was highly significant for all genes, indicating that the neo-squamous mucosa harbors a very greatly reduced malignant risk compared to untreated Barrett's disease. This is consistent with the normal histopathological appearance of the neo-squamous mucosa and the reassuring results of clinical studies regarding the long-term cancer risk after endoscopic therapy. The neo-squamous mucosa was significantly different to the normal control squamous mucosa for most genes but the significance of this finding is uncertain. One interpretation is that it suggests that attention should be given to careful inspection of the neo-squamous mucosa as well as, of course, any persistent BE areas after endotherapy. This could include taking random biopsies from a normal appearing neo-squamous mucosa; although the benefit of this is disputed it can rarely uncover buried (sub-squamous) BE or even adenocarcinoma.[17,33–35] Altogether, we interpret our results as providing some support for long-term endoscopic surveillance after endoscopic treatment of BE/IMC, even if the BE has been completely eradicated, we acknowledge that more extensive studies with longer follow-up periods are needed to more thoroughly evaluate the neo-squamous mucosa.

Conflict of interest

The authors declare that there is no conflict of interest.

Funding

This work was supported by the Australian National Health and Medical Research Council (project grant 100837) and the Cancer Council New South Wales (SRP 08-04).

References

1. Shaheen NJ, Falk GW, Iyer PG, et al. ACG Clinical guideline: Diagnosis and management of Barrett's esophagus. *Am J Gastroenterol* 2016; 111: 30–50.
2. Bahin FF, Jayanna M, Hourigan LF, et al. Long-term outcomes of a primary complete endoscopic resection strategy for short-segment Barrett's esophagus with high-grade dysplasia and/or early esophageal adenocarcinoma. *Gastrointest Endosc* 2016; 83: 68–77.
3. Pech O, May A, Manner H, et al. Long-term efficacy and safety of endoscopic resection for patients with mucosal adenocarcinoma of the esophagus. *Gastroenterology* 2014; 146: 652–660.
4. Whiteman DC, Appleyard M, Bahin FF, et al. Australian clinical practice guidelines for the diagnosis and management of Barrett's esophagus and early esophageal adenocarcinoma. *J Gastroenterol Hepatol* 2015; 30: 804–820.
5. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* 2014; 63: 7–42.
6. Bennett C, Moayyedi P, Corley DA, et al. BOB CAT: A large-scale review and delphi consensus for management of Barrett's esophagus with no dysplasia, indefinite for, or low-grade dysplasia. *Am J Gastroenterol* 2015; 110: 662–682.
7. Phoa KN, van Vilsteren FG, Weusten BL, et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: A randomized clinical trial. *JAMA* 2014; 311: 1209–1217.
8. Shaheen NJ, Overholt BF, Sampliner RE, et al. Durability of radiofrequency ablation in Barrett's esophagus with dysplasia. *Gastroenterology* 2011; 141: 460–468.

9. Phoa KN, Pouw RE, Bisschops R, et al. Multimodality endoscopic eradication for neoplastic Barrett oesophagus: Results of an European multicentre study (EURO-II). *Gut* 2015; 65(4): 555–562.
10. Moss A, Bourke MJ, Hourigan LF, et al. Endoscopic resection for Barrett's high-grade dysplasia and early esophageal adenocarcinoma: An essential staging procedure with long-term therapeutic benefit. *Am J Gastroenterol* 2010; 105: 1276–1283.
11. Wani S, Puli SR, Shaheen NJ, et al. Esophageal adenocarcinoma in Barrett's esophagus after endoscopic ablative therapy: A meta-analysis and systematic review. *Am J Gastroenterol* 2009; 104: 502–513.
12. Gupta M, Iyer PG, Lutzke L, et al. Recurrence of esophageal intestinal metaplasia after endoscopic mucosal resection and radiofrequency ablation of Barrett's esophagus: Results from a US multicenter consortium. *Gastroenterology* 2013; 145: 79–86.
13. Singh A and Chak A. Advances in the management of Barrett's esophagus and early esophageal adenocarcinoma. *Gastroenterol Rep (Oxf)* 2015; 3: 303–315.
14. Botelho NK, Schneiders FI, Lord SJ, et al. Gene expression alterations in formalin-fixed, paraffin-embedded Barrett esophagus and esophageal adenocarcinoma tissues. *Cancer Biol Ther* 2010; 10: 172–179.
15. Brabender J, Marjoram P, Salonga D, et al. A multigene expression panel for the molecular diagnosis of Barrett's esophagus and Barrett's adenocarcinoma of the esophagus. *Oncogene* 2004; 23: 4780–4788.
16. Chisholm JA, Mayne GC, Hussey DJ, et al. Molecular biomarkers and ablative therapies for Barrett's esophagus. *Expert Rev Gastroenterol Hepatol* 2012; 6: 567–581.
17. Pouw RE, Gondrie JJ, Rygiel AM, et al. Properties of the neosquamous epithelium after radiofrequency ablation of Barrett's esophagus containing neoplasia. *Am J Gastroenterol* 2009; 104: 1366–1373.
18. Krishnan K, Komanduri S, Cluley J, et al. Radiofrequency ablation for dysplasia in Barrett's esophagus restores beta-catenin activation within esophageal progenitor cells. *Dig Dis Sci* 2012; 57: 294–302.
19. Hage M, Siersema PD, Vissers KJ, et al. Molecular evaluation of ablative therapy of Barrett's oesophagus. *J Pathol* 2005; 205: 57–64.
20. Hage M, Siersema PD, Vissers KJ, et al. Genomic analysis of Barrett's esophagus after ablative therapy: Persistence of genetic alterations at tumor suppressor loci. *Int J Cancer* 2006; 118: 155–160.
21. Krishnadath KK, Wang KK, Taniguchi K, et al. Persistent genetic abnormalities in Barrett's esophagus after photodynamic therapy. *Gastroenterology* 2000; 119: 624–630.
22. Brabender J, Marjoram P, Lord RV, et al. The molecular signature of normal squamous esophageal epithelium identifies the presence of a field effect and can discriminate between patients with Barrett's esophagus and patients with Barrett's-associated adenocarcinoma. *Cancer Epidemiol Biomarkers Prev* 2005; 14: 2113–2117.
23. Lewis CJ, Thrumurthy SG, Pritchard S, et al. Comparison of COX-2, Ki-67, and BCL-2 expression in normal esophageal mucosa, Barrett's esophagus, dysplasia, and adenocarcinoma with postablation mucosa and implications for ablative therapies. *Surg Endosc* 2011; 25: 2564–2569.
24. Dijckmeester WA, Wijnhoven BP, Watson DI, et al. MicroRNA-143 and -205 expression in neosquamous esophageal epithelium following argon plasma ablation of Barrett's esophagus. *J Gastrointest Surg* 2009; 13: 846–853.
25. Phoa KN, Pouw RE, van Vilsteren FGI, et al. Remission of Barrett's esophagus with early neoplasia 5 years after radiofrequency ablation with endoscopic resection: A Netherlands cohort study. *Gastroenterology* 2013; 145: 96–104.
26. Orman ES, Kim HP, Bulsiewicz WJ, et al. Intestinal metaplasia recurs infrequently in patients successfully treated for Barrett's esophagus with radiofrequency ablation. *Am J Gastroenterol* 2013; 108: 187–195.
27. Krishnan K, Pandolfino JE, Kahrilas PJ, et al. Increased risk for persistent intestinal metaplasia in patients with Barrett's esophagus and uncontrolled reflux exposure before radiofrequency ablation. *Gastroenterology* 2012; 143: 576–581.
28. van Vilsteren F, Alvarez Herrero L, Pouw R, et al. Predictive factors for initial treatment response after circumferential radiofrequency ablation for Barrett's esophagus with early neoplasia: A prospective multicenter study. *Endoscopy* 2013; 45: 516–525.
29. Akiyama J, Marcus SN and Triadafilopoulos G. Effective intra-esophageal acid control is associated with improved radiofrequency ablation outcomes in Barrett's esophagus. *Dig Dis Sci* 2012; 57: 2625–2632.
30. Timmer MR, Brankley SM, Gorospe EC, et al. Prediction of response to endoscopic therapy of Barrett's dysplasia by using genetic biomarkers. *Gastrointest Endosc* 2014; 80: 984–991.
31. Prasad GA, Wang KK, Halling KC, et al. Utility of biomarkers in prediction of response to ablative therapy in Barrett's esophagus. *Gastroenterology* 2008; 135: 370–379.
32. Lord RV, Tsai PI, Danenberg KD, et al. Retinoic acid receptor-alpha messenger RNA expression is increased and retinoic acid receptor-gamma expression is decreased in Barrett's intestinal metaplasia, dysplasia, adenocarcinoma sequence. *Surgery* 2001; 129: 267–276.
33. Vaccaro BJ, Gonzalez S, Poneros JM, et al. Detection of intestinal metaplasia after successful eradication of Barrett's Esophagus with radiofrequency ablation. *Dig Dis Sci* 2011; 56: 1996–2000.
34. Shaheen NJ, Peery AF, Overholt BF, et al. Biopsy depth after radiofrequency ablation of dysplastic Barrett's esophagus. *Gastrointest Endosc* 2010; 72: 490–496.
35. Titi M, Overhiser A, Ulusarac O, et al. Development of subsquamous high-grade dysplasia and adenocarcinoma after successful radiofrequency ablation of Barrett's esophagus. *Gastroenterology* 2012; 143: 564–566.

Ann Surg Oncol (2015) 22:2431–2438
DOI 10.1245/s10434-014-4155-y

Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ORIGINAL ARTICLE – TRANSLATIONAL RESEARCH AND BIOMARKERS

High Expression of Cathepsin E in Tissues but Not Blood of Patients with Barrett's Esophagus and Adenocarcinoma

Oliver M. Fisher, MD[1], Angelique J. Levert-Mignon, BSc[1], Sarah J. Lord, MBBS, MS(Epi)[1,2,3], Natalia K. Botelho, BSc[1], Araluen K. Freeman, BSc[1], Melissa L. Thomas, BSc[1], Dan Falkenback, MD, PhD[1,4], Antony Wettstein, MBBS, FRACP[5], David C. Whiteman, BMedSc, MBBS, PhD, FAFPHM[6], Yuri V. Bobryshev, MD, PhD[1,7], and Reginald V. Lord, MBBS, MD, FRACS[1,8]

[1]St. Vincent's Centre for Applied Medical Research, Sydney, Australia; [2]NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia; [3]Department of Epidemiology and Medical Statistics, School of Medicine, University of Notre Dame, Sydney, Australia; [4]Department of Surgery, Lund University Hospital (Skane University Hospital) and Lund University, Lund, Sweden; [5]Diagnostic Endoscopy Centre, St. Vincent's Clinic, Sydney, Australia; [6]QIMR Berghofer Medical Research Institute, Brisbane, Australia; [7]Faculty of Medicine, School of Medical Sciences, University of New South Wales, Sydney, Australia; [8]Department of Surgery, School of Medicine, University of Notre Dame, Sydney, Australia

ABSTRACT

Background. Cathepsin E (CTSE), an aspartic proteinase, is differentially expressed in the metaplasia–dysplasia–neoplasia sequence of gastric and colon cancer. We evaluated CTSE in Barrett's esophagus (BE) and cancer because increased CTSE levels are linked to improved survival in several cancers, and other cathepsins are up-regulated in BE and esophageal adenocarcinoma (EAC).

Methods. A total of 273 pretreatment tissues from 199 patients were analyzed [31 normal squamous esophagus (NE), 29 BE intestinal metaplasia, 31 BE with dysplasia (BE/D), 108 EAC]. CTSE relative mRNA expression was measured by real-time polymerase chain reaction, and protein expression was measured by immunohistochemistry. CTSE serum levels were determined by enzyme-linked immunosorbent assay.

Results. Median CTSE mRNA expression levels were $\geq$1,000-fold higher in BE/intestinal metaplasia and BE/D compared to NE. CTSE levels were significantly lower in EAC compared to BE/intestinal metaplasia and BE/D, but significantly higher than NE levels. A similar expression pattern was present in immunohistochemistry, with absent staining in NE, intense staining in intestinal metaplasia and dysplasia, and less intense EAC staining. CTSE serum analysis did not discriminate patient groups. In a uni- and multivariable Cox proportional hazards model, CTSE expression was not significantly associated with survival in patients with EAC, although CTSE expression above the 25th percentile was associated with a 41 % relative risk reduction for death (hazard ratio 0.59, 95 % confidence interval 0.27–1.26, $p = 0.17$).

Conclusions. CTSE mRNA expression is up-regulated more than any known gene in Barrett intestinal metaplasia and dysplasia tissues. Protein expression is similarly highly intense in intestinal metaplasia and dysplasia tissues.

Barrett's esophagus (BE) is the condition in which the normal distal squamous lining of the esophagus is replaced by specialized metaplastic columnar epithelium.[1] BE is the strongest recognized risk factor for esophageal adenocarcinoma (EAC), a highly malignant cancer with an unparalleled 6-fold increase in incidence over the past three decades.[2] Less than 5 % of patients presenting with EAC have a previous diagnosis of BE because they have not undergone endoscopy, but even for patients under surveillance, there are significant problems, including sampling error and difficulties with the histopathologic interpretation

Electronic supplementary material The online version of this article (doi:10.1245/s10434-014-4155-y) contains supplementary material, which is available to authorized users.

First Received: 13 June 2014;
Published Online: 28 October 2014

R. V. Lord, MBBS, MD, FRACS
e-mail: rvlord@stvincents.com.au

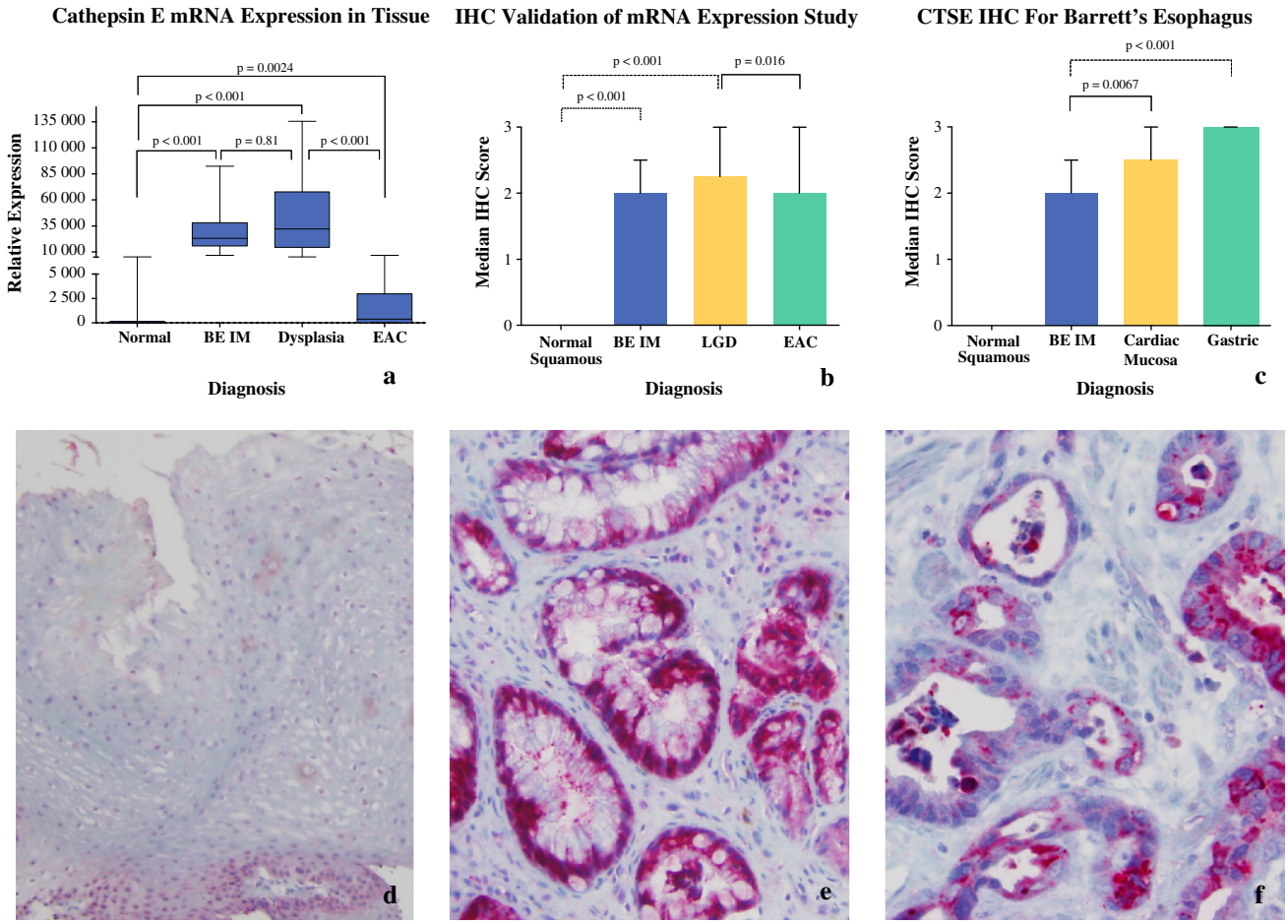


FIG. 1 CTSE mRNA expression analysis (**a**). CTSE is increased in BE and BE/D in comparison to NE. CTSE mRNA expression of CTSE in EAC is significantly lower than BE and dysplastic BE, but is significantly higher compared to NE. Immunohistochemical staining for CTSE (**d**–**f**). Marked, intense staining in BE and EAC with decreased staining intensity and shift in staining pattern with neoplastic progression of Barrett epithelium. Quantification of these findings could be statistically confirmed (**b**, **c**)

of degree of dysplasia, and EAC may develop between endoscopies.[3–5]

These problems have stimulated the search for clinically relevant biological markers, but so far, no biomarkers have proven sufficient for routine clinical practice.[6,7] Because of the problems with endoscopic diagnosis and surveillance, it is worth exploring non endoscopic, cheaper, and less-invasive diagnostic and monitoring options, such as blood, saliva, and brush cytology tests.

Cathepsin E (CTSE) is an intracellular aspartic protease that is normally expressed in a wide range of immune cells but also is present in osteoclasts and gastric epithelial cells; secreted forms have been described.[8–10] A differential expression pattern has been demonstrated for CTSE in normal, metaplastic, dysplastic, and neoplastic gastric epithelium as well as in the intestinal dysplasia–neoplasia sequence in $APC^{min/+}$ mice.[11–15] Furthermore, CTSE has also been suggested as both a diagnostic and a prognostic biomarker for some cancers.[16]

Other members of the cathepsin family (cathepsin B, C, D, K, and S) have been found to be up-regulated in BE and EAC, but analysis of CTSE in BE and EAC has not been reported.[17–20]

This study aimed to evaluate the potential prognostic value of CTSE to predict progression to more advanced disease in patients with Barrett metaplasia–dysplasia–adenocarcinoma spectrum, and to predict survival for patients with EAC.

MATERIALS AND METHODS

Study Design, Study Population, and Specimen Collection

A diagnostic case control analysis was performed to examine the associations among the following: (1) CTSE tissue mRNA expression and normal squamous esophagus (NE), BE, BE with dysplasia (BE/D) and EAC; (2) CTSE serum protein levels and NE, BE, BE/D and EAC; and (3)

the association between CTSE and overall EAC patient survival (Supplementary Fig. 1).

The NE, BE, and BE/D tissues as well as blood samples were collected at endoscopies performed at St. Vincent's Hospital, Sydney, from patients prospectively enrolled onto a research collaboration entitled PROBE-NET (Progression of Barrett's Esophagus to Esophageal Adenocarcinoma Network). The EAC specimens were obtained either from patients at St. Vincent's Hospital or from patients who had been enrolled onto the population-based case-control Australian Cancer Study.[21] All tissues were fixed in formalin and embedded in paraffin. The pathology diagnosis was established by pathologists at the respective host institutions. Before mRNA extraction, a section of each tissue sample was also sent for hematoxylin and eosin staining and reviewed to confirm the pathology in the research specimen. BE was defined as intestinal metaplasia with the presence of goblet cells. Patient serum samples were collected at study recruitment, centrifuged at 14,000× *g,* and then stored at −80 °C until further use.

For the analysis of CTSE as a prognostic marker for EAC survival, we used tissue samples from an independent cohort of 75 patients with early stage EAC (I–IIB) from the Australian Cancer Study.[21] All subjects had undergone treatment with potentially curative surgery alone and received no chemo- or radiotherapy. Patients who died within 30 days of surgery or who had cancer-involved operative resection margins (R1/R2 resection) were excluded.

Institutional review board approval for this study was obtained at all collaborating institutions, and all patients provided written informed consent.

RNA Isolation

From each paraffin-embedded tissue block, two 7 μm sections were cut and used for RNA extraction using the Qiagen FFPE RNeasy Kit (Cat. #74404; Qiagen, Valencia, CA) following the manufacturer's protocol. RNA yield and quality was measured using a Biospec Nano spectrophotometer (Shimadzu Scientific Instruments, Sydney, Australia).

Multiplexed Tandem Polymerase Chain Reaction

Multiplexed tandem polymerase chain reaction (MT-PCR) was used to quantitate the mRNA expression level of CTSE and a reference gene, *NONO* (non-POU domain containing, octamer-binding (NONO), transcript variant 2; NM_007363), using the Rotor-Gene 6000 real-time quantitative PCR system (Corbett Life Sciences/Qiagen, Sydney, Australia), as described previously.[22,23] Primers were designed with the help of Primer 3 software modified by AusDiagnostics Pty. Ltd. (AusDiagnostics, Alexandria, New South Wales, Australia), leading to a CTSE "inner" amplicon of 73 bp and an "outer" amplicon of 120 bp. Outer primer sequences for CTSE were 5′-CTCAATGGACCAGAGTGCCAAG-3′ (forward) and 5′-GAGGAGCCAGTGTCGAAGATG-3′ (reverse). Inner primer sequences were 5′-GAGTGCCAAGGAACCC CTCATC-3′ (forward) and 5′-TGGTGGGGAGCCAATGG AGATA-3′ (reverse). All primer pairs spanned an intron–exon boundary, and all samples were run in duplicate. The correct size and integrity of the products was verified on a Bioanalyzer DNA separation chip (Agilent Technologies, Forest Hill, Victoria, Australia).

CTSE Enzyme-Linked Immunosorbent Assay

A CTSE enzyme-linked immunosorbent assay (ELISA) kit (Biomatik Corp, Cambridge, Ontario, Canada) was used to measure CTSE protein levels in serum. Briefly, after reconstitution of all reagents, serum samples were incubated on precoated plates at 37 °C and 70 % humidity for 2 h. After addition of the primary antibody and incubation for another hour at 37 °C, plates were washed three times with wash buffer. Addition of the secondary antibody was followed by a further incubation for 30 min at 37 °C, and plates were then washed another five times before the addition of the reaction substrate. For antibody binding detection, the supplier's detection reagent was added for 15 min and the reaction halted by addition of the provided stopping solution. Plate readouts occurred in a 96-well multiplate reader (Multiskan Microplate Reader; Thermo Labsystems/Thermo Scientific, Waltham, MA) at an absorbance of 450 nm. All samples were assayed in triplicate and run without dilution. All plate readings had an intra-assay coefficient of variation <15 %.

Immunohistochemistry

Tissue specimens were processed in a standard fashion with regular formalin fixation and paraffin embedding. CTSE was identified in 5 μm tissue sections using a rabbit polyclonal anti-CTSE antibody (Cat. #ab36996; Abcam, Waterloo, NSW, Australia) in a standard alkaline phosphatase anti-alkaline phosphatase technique, as described previously.[24]

Immunohistochemistry Scoring

The sections were scored using a four-step scale: (0) no staining or equal to background, (1) weak diffuse cytoplasmic staining, (2) moderate cytoplasmic staining in at least 10 % of cells, and (3) strong immunostaining in a majority of cells.[25] Immunohistochemistry sections were scored by two experienced investigators who were blinded to clinical information. In cases of disagreement, consensus was reached after reanalysis on a multiheaded microscope.

TABLE 1 Clinical and pathologic characteristics of included patients

Characteristic	All ($n = 199$)		BE progression study ($n = 91$)		Serum analysis ($n = 33$)		Prognostic biomarker in EAC ($n = 75$)	
	n	%	*n*	%	*n*	%	*n*	%
Sex								
Male	164	82.4	67	73.6	29	87.9	68	90.7
Female	35	18.6	24	26.4	4	12.1	7	9.3
Age, year, median (IQR)	63	(55–71)	62	(53–69)	63	(54–70)	67	(59–74)
Diagnosis								
Healthy controls/normal squamous	31	15.6	22	24.2	9	27.3	–	–
BE intestinal metaplasia	29	14.6	21	23.1	8	24.2	–	–
BE with dysplasia	31	15.6	22	24.2	9	27.3	–	–
Esophageal adenocarcinoma	108	54.2	26	28.5	7	21.2	75	100
TNM (AJCC, 7th edition)								
Tis	2	1.9	2	7.7	–	–	–	–
T1–2	88	81.5	8	30.7	5	71.4	75	100
T3–4	8	7.4	6	23.1	2	28.6	–	–
N1–3	34	31.5	7	26.9	3	42.9	24	32.0
M+	3	2.8	3	11.5	–	–	–	–
Unknown T, N or M	10[a]	9.3	10[a]	38.5	–	–	–	–
Tumor stage (AJCC, 7th edition)								
0 (Tis)	2	1.9	2	7.7	–	–	–	–
IA–B	33	30.6	2	7.7	3	42.9	28	37.3
IIA	26	24.1	2	7.7	1	14.3	23	30.7
IIB	29	26.9	4	15.4	1	14.3	24	32.0
IIIA–C	7	6.5	5	19.2	2	28.6	–	–
IV	3	2.8	3	11.5	–	–	–	–
Unknown stage	8[a]	7.4	8[a]	30.8	–	–	–	–
Survival, d, median (range)	1,182	(630–1,685)	1,305	(195–1,780)	1,277	(766–1,342)	1,161	(724–1,663)

Totals may not equal 100 % due to rounding

BE Barrett's esophagus, *EAC* esophageal adenocarcinoma, *IQR* interquartile range, *TNM* tumor, node, metastasis classification system, *AJCC* American Joint Committee on Cancer

[a] Two patients included had no clinical data on T and N status but were found to be M+ at assessment. Regardless of this, primary tissue samples were used for analysis in the respective study. These patients were excluded from survival analysis

Statistical Analysis

The mRNA raw expression values were obtained on the Rotor-Gene MT-PCR system, and then relative expression values were calculated as the ratio of the mRNA level of CTSE to the control gene *NONO,* with the expression of *NONO* set to a fixed level (1000). Where necessary, $\log_2$ transformation of relative expression values and/or serum values was performed to achieve normal distribution. Differences between two groups were measured by Student's *t* test or the Wilcoxon rank-sum test. One-way analysis of variance was used to compare differential gene and protein expression between patient groups. The Kaplan–Meier method was used for survival estimates, and differences in survival were analyzed using the log-rank test. Cox proportional hazards models were used for uni- and multivariable analysis. All *p* values of ≤ 0.05 were regarded as statistically significant. All analyses were performed by the SAS Statistical Package, version 9.2 (SAS Institute, Cary, NC). Prism (GraphPad Prism version 6.0c for Mac OS X; GraphPad Software, San Diego, CA) was used for graphs.

RESULTS

Patients and Tissues

As shown in Table 1, a total of 273 tissue specimens from 199 patients were included. Ninety-one patients were studied to evaluate CTSE as a marker for the progression of BE to EAC, 33 patients provided serum samples to evaluate CTSE as a biomarker in blood, and 75 early-stage EAC patients were included in the evaluation of CTSE as a prognostic biomarker.

Each part of the study included an independent cohort of patients, thus allowing for intrastudy validation of CTSE as a marker for the respective pathologies. Despite chart review, the correct tumor stage could not be assessed in 8 patients (7.4 %) as a result of incomplete clinical data.

Expression Analysis As shown in Fig. 1a, median CTSE mRNA relative expression was more than 1,000-fold higher in BE compared to NE (18.41 vs. 23,221; $p < 0.001$). Median CTSE mRNA expression in EAC was lower than in BE and dysplastic BE ($p < 0.001$) but higher than in NE (875.14 vs. 18.41; $p = 0.0024$).

Immunohistochemistry All BE specimens and all EAC specimens stained strongly for CTSE, with high specificity to the glandular structures and almost absent staining of the stromal fraction of the esophageal specimens. Figure 1d–f provides representative immunostaining patterns of the respective histopathology tissue types.

CTSE staining was completely absent within the squamous epithelium, whereas in BE/D median staining scores were 2.25 ($p < 0.001$, Fig. 1b). Staining intensities in EAC were similar to BE (median staining score 2.0), but the location shifted more apically and the staining pattern showed more granular features in EAC. Staining in EAC was significantly lower than in BE/D (2.0 vs. 2.25; $p = 0.016$, Fig. 1b).

CTSE immunostaining was also assessed in cardiac and gastric fundus mucosa because CTSE is known to be present in gastric glands.[13] CTSE staining was significantly higher in both proximal gastric mucosae compared to BE (2.0 vs. 2.5 and 3.0; $p = 0.0067$ and <0.001, respectively; Fig. 1c).

CTSE Serum ELISA As shown in Fig. 2, there were no significant differences between patient groups.

Analysis of CTSE as a Prognostic Biomarker in Early-Stage EAC

Patient Survival Overall median survival of the patients in the independent EAC cohort was 3.2 years (38.7 months), and overall 5-year survival was 66 %.

T1a and T1b patients showed a significantly increased survival (58.7 and 46.0 months, respectively) compared to T2 patients (25.2 months; $p < 0.001$). Stage I (IA+B) patients had a median survival of 41.2 months (3.4 years), whereas stage II (IIA+IIB) survived 31.4 months (2.61 years, $p = 0.0027$).

CTSE mRNA Expression and Tumor Stage

No significant difference was found in CTSE expression levels between T stages, American Joint Committee on Cancer (AJCC) stage I versus II cancers, lymph node negative versus positive disease, or male versus female sex (data not shown).

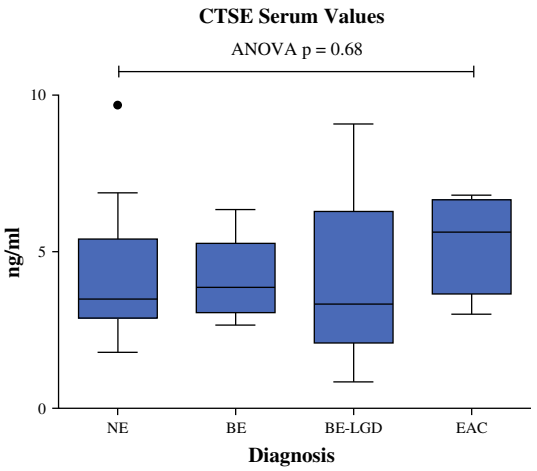


FIG. 2 CTSE serum values. Serum values do not differ between pathologic patient groups, although tissue levels are markedly increased

CTSE EAC Tissue mRNA Expression and Survival For survival analysis, CTSE expression values were dichotomized at the 25th percentile, median, and 75th percentile to determine the influence of CTSE expression on overall patient survival. Kaplan–Meier survival curves showed that patients with a CTSE expression above the 25th percentile had a non-significant trend toward improved overall survival (log-rank $p = 0.14$, Fig. 3). In uni- and multivariable analysis, elevated CTSE expression levels were not significantly associated with survival [hazard ratio (HR) 0.65; 95 % confidence interval (CI) 0.73–3.24, $p = 0.25$). CTSE expression above the 25th percentile was associated with a non-significant 41 % relative risk reduction for death (HR 0.59, 95 % CI 0.27–1.26, $p = 0.17$). In a backward stepwise regression model including sex, age, overall tumor stage, and CTSE expression below the 25th percentile, only age (HR 1.04, 95 % CI 1.00–1.08; $p = 0.04$) and AJCC stage II (HR 4.93, 95 % CI 1.88–12.88; $p = 0.001$) were independent prognostic markers for decreased survival (Table 2).

DISCUSSION

This novel study shows that CTSE is highly overexpressed in BE and BE/D compared to normal esophageal tissue. CTSE mRNA expression was 1,000-fold higher in BE compared to normal esophageal tissue, which we believe to be the highest gene expression change reported for this disease. Lower levels of CTSE mRNA were observed in EAC compared to BE. A similar expression pattern has been demonstrated for CTSE in

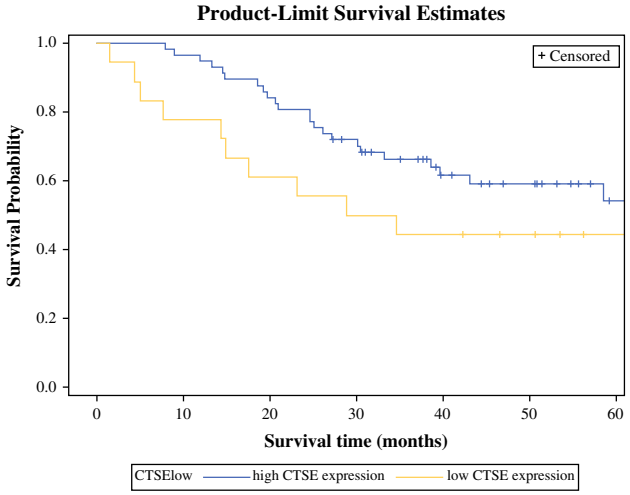


FIG. 3 CTSE mRNA expression and survival. Dichotomization at log_2-transformed 25th percentile. Patients with CTSE over the 25th percentile show a 41 % reduction in risk for death (HR 0.59, 95 % CI 0.27–1.26; $p = 0.17$)

TABLE 2 Uni- and multivariable analysis for factors contributing to mortality according to the Cox proportional hazards model

Variable	Univariable analysis			Multivariable analysis		
	HR	95 % CI	*p*	HR	95 % CI	*p*
Age	1.04	1.00–1.08	0.04	1.04	1.00–1.08	0.04
Sex (male)	1.71	0.41–7.15	0.464	1.49	0.35–6.38	0.59
AJCC stage II	4.68	1.80–12.14	0.002	4.93	1.88–12.88	0.001
Log CTSE <25th percentile	1.54	0.73–3.24	0.25	1.71	0.79–3.65	0.17
Log CTSE >25th percentile	0.65	0.31–1.36	0.25	0.59	0.27–1.26	0.17

HR hazard ratio, *CI* 95 % confidence interval, *AJCC* American Joint Committee on Cancer, *CTSE* cathepsin E

normal, metaplastic, dysplastic, and neoplastic gastric epithelium as well as in the intestinal dysplasia–neoplasia sequence in $APC^{min/+}$ mice.[11–15] One explanation for our observed CTSE overexpression is that exposure to gastric refluxate induces expression of gastric proteases; consistent with this we found that despite the remarkable induction in BE tissues, CTSE levels are still lower than those found in either gastric fundus or cardiac mucosa.

Alternatively, the remarkably high CTSE expression levels may indicate a functional role in this disease. Unfortunately, the exact function of CTSE remains to be defined because the specific substrate for this protease is not known.[10,26] A role in host defense has been suggested because of the high CTSE expression in immune and antigen presenting cells, but CTSE was solely expressed in glandular BE cells and not in stromal cells in the present study, similar to the distribution observed in a mouse intestinal neoplasia study.[14]

Interestingly, the other cathepsin family members cathepsin B, C, K, and S are also up-regulated in BE and EAC, and cathepsin D (CTSD) mRNA expression shows a significant stepwise increase in erosive esophagitis, intestinal metaplasia and EAC.[17–20] Further functional hypotheses for CTSE in Barrett disease involve the contents of the gastro-esophageal refluxate. Intracellular and secreted CTSD requires a low pH to exert its proteolytic activity, leading to the speculation that CTSD activity may be especially enhanced in the acidic environment of gastroesophageal reflux associated disease.[20,27] CTSD is also involved in the resistance to the bile salt deoxycholate–induced apoptosis in colon cancer cell lines.[28] Because CTSE is highly homologous to CTSD, there may be a similar acid and bile-associated function for CTSE in the context of BE development.[9,26]

A conclusive explanation for the significantly lower CTSE expression levels in EAC compared to BE is not available. In particular, it is not clear if CTSE expression is down-regulated in EAC and thus CTSE more a marker of BE than of EAC. In other cancers, CTSE has been shown to exert an antitumorigenic effect in prostate cancer cells—for example, by acting as the cleavage enzyme for tumor necrosis factor-related apoptosis ligand (TRAIL), which has also been implicated in the pathogenesis of EAC.[29,30] Injection of purified CTSE into human tumor xenografts results in a dose-dependent induction of apoptosis and inhibition of tumor growth.[29] It can therefore be speculated that the increased levels of CTSE in BE and BE/D may serve a protective mechanism. In this hypothesis, the down-regulation of CTSE marks an enhanced susceptibility to neoplasia formation, as suggested by a mouse melanoma study.[29] Further, loss of CTSE expression has also been shown to induce mammary gland neoplasia.[31]

High levels of CTSE have been shown to be associated with improved survival in various cancers, but in EAC we found only a non-significant trend in which expression of CTSE above the 25th percentile resulted in a 41 % risk reduction for death.[31–33]A larger study including patients with worse disease stage could be undertaken, as limited power and small survival differences due to the inclusion of only early stage, chemoradiotherapy-naive patients may have reduced our ability to detect a statistically significant association.

Finally, although highly desirable from a clinical perspective, this study indicates a lack of value in measuring CTSE protein levels in serum. Although there was a non-significant trend to higher CTSE protein levels in patients with EAC, our exploratory study was not powered to detect small differences in CTSE expression between patient groups. Alternatively, however, CTSE activity levels could be studied according to a recent report, which showed that CTSE activity levels but not protein levels were associated

with more advanced disease, recurrence, and prognosis in patients with breast cancer.[31]

CONCLUSIONS

The remarkable induction of CTSE expression in BE intestinal metaplasia and dysplasia, together with the significant down-regulation in EAC tissues, suggests a possible role for CTSE in the Barrett disease spectrum. The intense CTSE protein expression in BE and lower levels of expression in EAC could be evaluated by pathologists as a method to simplify the evaluation of esophageal tissues, although we acknowledge that further studies are required to substantiate this potential benefit.

ACKNOWLEDGMENT This work was supported by the National Health and Medical Research Council (NHMRC 1040947), Cancer Council New South Wales (SRP 08-04), and St. Vincent's Clinic Foundation, Sydney. O.M.F. is supported by the Swiss Cancer League (BIL KLS-3133-02-2013). D.F. is supported by the Swedish Society of Medicine, the Maggie Stephens Foundation, and the Sparre Foundation.

DISCLOSURE The authors declare no conflict of interest.

REFERENCES

1. Phillips WA, Lord RV, Nancarrow DJ, Watson DI, Whiteman DC. Barrett's esophagus. *J Gastroenterol Hepatol.* 2011;26:639–648.
2. Eheman C, Henley SJ, Ballard-Barbash R, et al. Annual report to the nation on the status of cancer, 1975–2008, featuring cancers associated with excess weight and lack of sufficient physical activity. *Cancer.* 2012;118:2338–2366.
3. Dulai GS, Guha S, Kahn KL, Gornbein J, Weinstein WM. Preoperative prevalence of Barrett's esophagus in esophageal adenocarcinoma: a systematic review. *Gastroenterology.* 2002;122:26–33.
4. Spechler SJ. Screening and surveillance for Barrett's esophagus—an unresolved dilemma. Nature clinical practice. *Gastroenterol Hepatol.* 2007;4:470–471.
5. Reid BJ, Haggitt RC, Rubin CE, et al. Observer variation in the diagnosis of dysplasia in Barrett's esophagus. *Hum Pathol.* 1988;19:166–178.
6. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut.* 2014;63:7–42.
7. Varghese S, Lao-Sirieix P, Fitzgerald RC. Identification and clinical implementation of biomarkers for Barrett's esophagus. *Gastroenterology.* 2012;142(3):435–441. e2.
8. Yoshimine Y, Tsukuba T, Isobe R, et al. Specific immunocytochemical localization of cathepsin E at the ruffled border membrane of active osteoclasts. *Cell Tissue Res.* 1995;281:85–91.
9. Kageyama T, Takahashi K. A cathepsin D–like acid proteinase from human gastric mucosa. Purification and characterization. *J Biochem.* 1980;87:725–735.
10. Yamamoto K, Kawakubo T, Yasukochi A, Tsukuba T. Emerging roles of cathepsin E in host defense mechanisms. *Biochim Biophys Acta.* 2012;1824:105–112.
11. Saku T, Sakai H, Tsuda N, et al. Cathepsins D and E in normal, metaplastic, dysplastic, and carcinomatous gastric tissue: an immunohistochemical study. *Gut.* 1990;31:1250–1255.
12. Konno-Shimizu M, Yamamichi N, Inada K, et al. Cathepsin E is a marker of gastric differentiation and signet-ring cell carcinoma of stomach: a novel suggestion on gastric tumorigenesis. *PLoS One.* 2013;8:e56766.
13. Matsuo K, Kobayashi I, Tsukuba T, et al. Immunohistochemical localization of cathepsins D and E in human gastric cancer: a possible correlation with local invasive and metastatic activities of carcinoma cells. *Hum Pathol.* 1996;27:184–190.
14. Busquets L, Guillen H, DeFord ME, et al. Cathepsin E is a specific marker of dysplasia in APC mouse intestine. *Tumour Biol.* 2006;27:36–42.
15. Paoni NF, Feldman MW, Gutierrez LS, Ploplis VA, Castellino FJ. Transcriptional profiling of the transition from normal intestinal epithelia to adenomas and carcinomas in the $APC^{Min/+}$ mouse. *Physiol Genomics.* 2003;15:228–235.
16. Uno K, Azuma T, Nakajima M, et al. Clinical significance of cathepsin E in pancreatic juice in the diagnosis of pancreatic ductal adenocarcinoma. *J Gastroenterol Hepatol.* 2000;15:1333–1338.
17. Hughes SJ, Glover TW, Zhu XX, et al. A novel amplicon at 8p22–23 results in overexpression of cathepsin B in esophageal adenocarcinoma. *Proc Natl Acad Sci USA.* 1998;95:12410–12415.
18. Cheng P, Gong J, Wang T, et al. Gene expression in rats with Barrett's esophagus and esophageal adenocarcinoma induced by gastroduodenoesophageal reflux. *World J Gastroenterol.* 2005;11:5117–122.
19. Luthra MG, Ajani JA, Izzo J, et al. Decreased expression of gene cluster at chromosome 1q21 defines molecular subgroups of chemoradiotherapy response in esophageal cancers. *Clin Cancer Res.* 2007;13:912–929.
20. Breton J, Gage MC, Hay AW, et al. Proteomic screening of a cell line model of esophageal carcinogenesis identifies cathepsin D and aldo-keto reductase 1C2 and 1B10 dysregulation in Barrett's esophagus and esophageal adenocarcinoma. *J Proteome Res.* 2008;7:1953–1962.
21. Whiteman DC, Sadeghi S, Pandeya N, et al. Combined effects of obesity, acid reflux and smoking on the risk of adenocarcinomas of the oesophagus. *Gut.* 2008;57:173–180.
22. Stanley KK, Szewczuk E. Multiplexed tandem PCR: gene profiling from small amounts of RNA using SYBR Green detection. *Nucleic Acids Res.* 2005;33:e180.
23. Botelho NK, Schneiders FI, Lord SJ, et al. Gene expression alterations in formalin-fixed, paraffin-embedded Barrett esophagus and esophageal adenocarcinoma tissues. *Cancer Biol Ther.* 2010;10:172–179.
24. Bobryshev YV, Freeman AK, Botelho NK, et al. Expression of the putative stem cell marker Musashi-1 in Barrett's esophagus and esophageal adenocarcinoma. *Dis Esophagus.* 2010;23:580–589.
25. Buskens CJ, Van Rees BP, Sivula A, et al. Prognostic significance of elevated cyclooxygenase 2 expression in patients with adenocarcinoma of the esophagus. *Gastroenterology.* 2002;122:1800–1807.
26. Zaidi N, Hermann C, Herrmann T, Kalbacher H. Emerging functional roles of cathepsin E. *Biochem Biophys Res Commun.* 2008;377:327–330.
27. van der Stappen JW, Williams AC, Maciewicz RA, Paraskeva C. Activation of cathepsin B, secreted by a colorectal cancer cell

line requires low pH and is mediated by cathepsin D. *Int J Cancer.* 1996;67:547–554.
28. Bernstein H, Payne CM, Kunke K, et al. A proteomic study of resistance to deoxycholate-induced apoptosis. *Carcinogenesis.* 2004;25:681–692.
29. Kawakubo T, Okamoto K, Iwata J, et al. Cathepsin E prevents tumor growth and metastasis by catalyzing the proteolytic release of soluble TRAIL from tumor cell surface. *Cancer Res.* 2007;67:10869–10878.
30. Clemons N, Phillips W, Lord RV. Signaling pathways in the molecular pathogenesis of adenocarcinomas of the esophagus and gastresophageal junction. *Cancer Biol Ther.* 2013;14(9): 782–795.
31. Kawakubo T, Yasukochi A, Toyama T, et al. Repression of cathepsin E expression increases the risk of mammary carcinogenesis and links to poor prognosis in breast cancer. *Carcinogenesis.* 2014; 35(3):714–726. In press.
32. Ullmann R, Morbini P, Halbwedl I, et al. Protein expression profiles in adenocarcinomas and squamous cell carcinomas of the lung generated using tissue microarrays. *J Pathol.* 2004;203:798–807.
33. Fristrup N, Ulhoi BP, Birkenkamp-Demtroder K, et al. Cathepsin E, maspin, Plk1, and survivin are promising prognostic protein markers for progression in non-muscle invasive bladder cancer. *Am J Pathol.* 2012;180:1824–1834.

JGH Journal of Gastroenterology and Hepatology

JGHF

doi:10.1111/jgh.12913

SOLICITED REVIEW

Australian clinical practice guidelines for the diagnosis and management of Barrett's esophagus and early esophageal adenocarcinoma

David C Whiteman,* Mark Appleyard,† Farzan F Bahin,‡,§ Yuri V Bobryshev,¶,** Michael J Bourke,‡,§ Ian Brown,†† Adrian Chung,‡‡ Andrew Clouston,†† Emma Dickins,§§ Jon Emery,¶¶ Guy D Eslick,*** Louisa G Gordon,††† Florian Grimpen,† Geoff Hebbard,‡‡‡ Laura Holliday,§§ Luke F Hourigan,§§§ Bradley J Kendall,*,§§§ Eric YT Lee,‡ Angelique Levert-Mignon,** Reginald V Lord,**,¶¶¶ Sarah J Lord,**,¶¶¶ Derek Maule,**** Alan Moss,¶¶,†††† Ian Norton,‡‡‡‡ Ian Olver,§§ Darren Pavey,§§§§ Spiro Raftopoulos,¶¶¶¶ Shan Rajendra,***** Mark Schoeman,†††††,‡‡‡‡‡ Rajvinder Singh,‡‡‡‡‡,§§§§§ Freddy Sitas,¶,***,**** B Mark Smithers,§§§ Andrew C Taylor,¶¶¶¶¶ Melissa L Thomas,**,¶¶¶ Iain Thomson,§§§ Henry To,****** Jutta von Dincklage,§§ Christine Vuletich,§§ David I Watson****** and Ian F Yusoff¶¶¶¶

*QIMR Berghofer Medical Research Institute, †Royal Brisbane and Women's Hospital, ††Envoi Pathology, †††Griffith University, §§§Princess Alexandra Hospital, University of Queensland, Brisbane, Queensland, ‡Westmead Hospital, §Westmead Clinical School, University of Sydney, ¶University of New South Wales, **St Vincent's Centre for Applied Medical Research, §§Cancer Council Australia, ***University of Sydney, ¶¶¶School of Medicine, University of Notre Dame Australia, ****Cancer Council New South Wales, ‡‡‡‡Royal North Shore Hospital, §§§§Bankstown and Concord Hospitals, *****Bankstown-Lidcombe Hospital, Sydney, New South Wales, ‡‡Flinders Medical Centre, †††††Royal Adelaide Hospital, ‡‡‡‡‡University of Adelaide, §§§§§The Lyell McEwin Hospital, Adelaide, South Australia, ¶¶University of Melbourne, ‡‡‡Royal Melbourne Hospital, ††††Western Health, ¶¶¶¶¶St Vincent's Hospital, ******Peter MacCallum Cancer Centre, Melbourne, Victoria, ¶¶¶¶Sir Charles Gairdner Hospital, Perth, Western Australia, Australia

Key words
Barrett's esophagus, clinical practice, esophageal adenocarcinoma, guidelines.

Accepted for publication 19 December 2014.

Correspondence
Professor David C Whiteman, Cancer Control Group, QIMR Berghofer Medical Research Institute, Locked Bag 2000, Royal Brisbane and Women's Hospital, Brisbane, Qld. 4029, Australia. Email: david.whiteman@qimrberghofer.edu.au

Abstract

Barrett's esophagus (BE), a common condition, is the only known precursor to esophageal adenocarcinoma (EAC). There is uncertainty about the best way to manage BE as most people with BE never develop EAC and most patients diagnosed with EAC have no preceding diagnosis of BE. Moreover, there have been recent advances in knowledge and practice about the management of BE and early EAC. To aid clinical decision making in this rapidly moving field, Cancer Council Australia convened an expert working party to identify pertinent clinical questions. The questions covered a wide range of topics including endoscopic and histological definitions of BE and early EAC; prevalence, incidence, natural history, and risk factors for BE; and methods for managing BE and early EAC. The latter considered modification of lifestyle factors; screening and surveillance strategies; and medical, endoscopic, and surgical interventions. To answer each question, the working party systematically reviewed the literature and developed a set of recommendations through consensus. Evidence underpinning each recommendation was rated according to quality and applicability.

Author contributions: Davic C Whiteman, Mark Appleyard, Farzan Fahrtash Bahin, Yuri V Bobryshev, Michael J Bourke, Ian Brown, Adrian Chung, Andrew Clouston, John Emery, Guy D Eslick, Louisa G Gordon, Florian Grimpen, Geoff Hebbard, Luke Hourigan, Bradley J Kendall, Eric Y T Lee, Angelique Levert, Reginald V Lord, Sarah J Lord, Alan Moss, Ian Norton, Darren Pavey, Spiro Raftopoulos, Shan Rajendra, Mark Schoeman, Rajvinder Singh, Freddy Sitas, Mark Smithers, Andrew Taylor, Melissa L Thomas, Iain Thomson, Henry To, David I Watson, and Ian F Yusoff reviewed the literature and compiled the evidence summaries. Emma Dickins and Laura Holliday conducted systematic literature searches, screened the primary literature, and collated the evidence summaries. Jutta von Dincklage and Christine Vuletich managed the guideline development process and provided project governance. Ian Olver provided oversight and funding and Derek Maule provided consumer input. All authors were involved in drafting and critical revision of the manuscript.

Introduction

Barrett's esophagus (BE) is the only known precursor to esophageal adenocarcinoma (EAC), a cancer with a rapidly rising incidence. Most people with BE never develop EAC however, and most patients diagnosed with EAC have no preceding diagnosis of BE. Thus, there is uncertainty about the best way to manage this condition.

These guidelines about BE and early EAC are aimed at gastroenterologists, pathologists, surgeons and physicians, and other members of multidisciplinary teams to which patients with BE and EAC are referred. The guidelines will also be relevant to primary care practitioners and patients diagnosed with this condition. The need to develop Australian guidelines for the management of BE and early EAC was identified as a priority by a strategic partnership of clinicians, researchers, patients, and policy makers initiated by Cancer Council NSW in 2011.

Information covered by the guidelines includes:

1. Endoscopic and histological definitions of BE and early EAC
2. Prevalence, incidence, natural history, and risk factors for BE
3. Management of BE and early EAC, including modification of lifestyle factors, screening, surveillance, and medical, endoscopic, and surgical interventions.

The evidence summaries and recommendations are provided separately for BE without dysplasia and BE with dysplasia and/or early cancer, but do not extend to the management of invasive EAC. The recommendations contained herein should not override good clinical judgment. However, they do represent consensus views of expert practitioners and accord with international practices. This publication represents a summary of more extensive material hosted on the Cancer Council Australia Wiki platform[1] that explores the reasons underlying the recommendations in more detail.

Methods

Guideline development was facilitated by Cancer Council Australia, which managed the project and provided in-kind support. No external funding was received for guideline development.

The guidelines were developed by a multidisciplinary working group and used standard methodology.[2] A series of clinical questions were developed to be answered based on systematic reviews. In consultation with the working group, systematic search strategies were developed by project officers using the PICO framework and limits and exclusion criteria were pre-defined to complete the systematic review protocol. Databases searched included the Cochrane Library, PubMed, Embase, Trip Database, Econlit, National Health Service (UK) Economic Evaluation Database, the National Guideline Clearinghouse, the National Comprehensive Cancer Network and the National Institute for Health and Clinical Excellence, Scottish Intercollegiate Guidelines Network, and Canadian Medical Association. Search results were screened by project officers and relevant articles were sent to topic authors for critical appraisal with respect to level and quality of evidence, effect size, and clinical importance and relevance. The level of evidence for each article was assigned according to the National Health and Medical Research Council of Australia Evidence Hierarchy (Table 1).

Table 1 Hierarchy of evidence recommendation†

Level	Description
I	A systematic review of level II studies
II	A randomized controlled trial (intervention) or a prospective cohort study (etiology)
III-1	A pseudo-randomized controlled trial (intervention) or all or none design (etiology)
III-2	A comparative study with concurrent controls (intervention) or a retrospective cohort study (etiology)
III-3	A comparative study without concurrent controls (intervention) or a case–control study (etiology)
IV	Case series with either post-test or pre-test/post-test outcomes or a cross-sectional study

†Adapted from the National Health and Medical Research Council of Australia.

Table 2 Body of evidence recommendation†

Grade	Description
A	Body of evidence can be trusted to guide practice
B	Body of evidence can be trusted to guide practice in most situations
C	Body of evidence provides some support for recommendation(s) but care should be taken in its application
D	Body of evidence is weak and recommendation must be applied with caution
Practice point	Where no good-quality evidence is available but there is consensus among expert working group members, so-called Practice points are given

†Adapted from the National Health and Medical Research Council of Australia.

Each topic author summarized the relevant body of literature and then developed recommendations. Each recommendation was assigned a grade by the working group taking into account the volume, consistency, generalizability, applicability, and clinical impact of the supporting evidence (Table 2). When there was insufficient evidence to make a specific recommendation but consensus among experts about the advisability of making a clinically relevant statement, the working group formulated "practice points" to guide clinical practice. The working group also reviewed comparable international guidelines to calibrate the recommendations.

The draft guidelines underwent public consultation in June and July 2014. Feedback was reviewed by topic authors and the working group. Subsequent changes to the draft were agreed by consensus of the working group and the final guidelines were released on August 2014. The Wiki guidelines will be reviewed annually and updated as required.

Guidelines for BE without dysplasia

What is the definition of BE and how is it described? BE is a premalignant condition of the esophagus defined as the presence of metaplastic columnar epithelium,[3]

which appears endoscopically as salmon pink mucosa extending above the gastro-esophageal junction (GEJ) and into the tubular esophagus, thereby replacing the normal stratified squamous epithelium.[3,4] An accurate diagnosis of BE depends on the endoscopic recognition of the anatomic landmarks at the GEJ and squamocolumnar junction.[5] Using the Prague C&M (circumferential and maximal) criteria proposed by the International Working Group for the Classification of Esophagitis,[6] the landmark for the GEJ is the proximal end of the gastric folds.

The metaplastic columnar mucosa can be one of three types: gastric-fundic type, cardiac type, and intestinal type.[7] There remains disagreement as to the histological features of the columnar mucosa necessary to define BE as reflected in the differing definitions given in European and American guidelines.[8–11] For the Australian guidelines, however, the presence of intestinal metaplasia with morphologically typical goblet cells was considered necessary for the diagnosis of BE.

Biopsies from the tubular esophagus containing columnar mucosa without intestinal metaplasia should be given a descriptive diagnosis (e.g. columnar mucosa without intestinal metaplasia), but it is currently recommended that these are not diagnosed as BE until the biological significance of this entity is clarified.

Intestinal metaplasia occurring in isolation at the GEJ or cardia without metaplasia in the tubular esophagus is not considered BE. It may be a precursor to carcinoma, but the risk is low and surveillance is not warranted.[12,13] However, goblet cells noted in a GEJ biopsy can be confirmed to be intestinal metaplasia in columnar-lined esophagus (CLE) if the particular biopsy fragment shows native esophageal structures such as submucosal glands and/or ducts.

Practice points. To identify patients at increased risk of neoplastic progression, BE is defined as metaplastic columnar mucosa in the tubular esophagus, with intestinal metaplasia proven histologically.

Biopsies to confirm intestinal metaplasia should be performed when any length of possible BE is seen extending above the GEJ.

The extent of BE should be described using the Prague C&M criteria.

What is the optimal tissue sampling at endoscopy for diagnosis of BE?

Intestinal metaplasia can be patchy and may not be consistently sampled with endoscopic biopsies[14] (level of evidence IV). Advancements in chromoendoscopy (methylene blue, indigo carmine, and acetic acid), endoscope digital enhancements (narrow-band imaging, i-SCAN, Fujinon intelligent chromo endoscopy), and enhanced magnification have not been shown to be superior to the currently accepted practice of random four-quadrant biopsies at 2-cm intervals[15–17] (levels of evidence I, II, IV, respectively); however, the diagnostic yield may be higher with increasing number of biopsies (level of evidence IV).[18] Jumbo biopsy forceps have not been shown to be superior to standard capacity forceps in obtaining adequate biopsy samples (level of evidence II).[19] Office-based unsedated transnasal endoscopy using pediatric biopsy forceps is well tolerated and may emerge as a cost-effective strategy (level of evidence II).[20–22]

Recommendation. Random four-quadrant biopsies at 2-cm intervals are the mainstay for tissue sampling (recommendation grade B).

Practice points. Focal abnormalities such as ulcerated or nodular lesions should be targeted with biopsies and labeled before random biopsies from the rest of the mucosa as minor biopsy-related bleeding is common and may impair endoscopic views.

Technological advancements in chromoendoscopy, digital enhancements, and enhanced magnification complement rather than replace random four-quadrant biopsies at 2-cm intervals. Biopsies obtained every 2 cm should be placed into separate jars that are labeled according to the distance from the incisors, while biopsies from the GEJ and cardia can also be specifically labeled as such.

Are there biomarkers for the diagnosis of BE?

Numerous biomarkers have been proposed to aid the diagnosis of BE. Estimates of diagnostic accuracy have been reported for tissue biomarkers, including cytokeratin profiling,[23–29] immunohistochemical biomarkers to detect goblet cells such as mucin immunostaining,[30,31] and stress response protein AG2;[32] a serum biomarker (G17[33]); and a non-endoscopic capsule sponge device to collect cytology samples for Trefoil factor 3 immunohistochemistry (TFF3)[34,35] (diagnostic accuracy level of evidence II–III-3). These studies provide insufficient evidence to recommend any biomarkers to supplement or replace standard practice use of endoscopy and histopathology due to study designs with a high risk of bias, wide variation in accuracy estimates across studies, and no comparison with current standard practice.

Recommendation. There is insufficient evidence to recommend cytokeratins, MUC, G17, or AG2 to aid BE diagnosis (grade D).

There is insufficient evidence to recommend the non-endoscopic capsule sponge device with TFF3 for BE screening (grade C).

What is the prevalence of BE in the Australian population in comparison with other populations?

Globally, the prevalence of BE is low (<5%) but is higher in selected groups such as those with gastro-esophageal reflux disease (>15%). There are no studies describing the prevalence of BE in an asymptomatic, unselected Australian population. One small study suggests a high prevalence in high-risk patient populations.[36] A data linkage study conducted in one Australian healthcare region reported prevalence rates at each of three time points as 0.42% (1990), 2.3% (1998), and 4.2% (2002).[37] International studies suggest prevalence varies significantly by ethnicity (e.g. Asians <1% prevalence) and gender (more common in males).

Which factors best predict the risk of developing BE?

Risk factors for BE have been assessed in more than 50 studies. All studies have been observational, and most have been case–control studies of variable quality. From these studies, the major risk factors identified include age,[38] male sex,[39] history of frequent gastro-esophageal acid reflux,[40] central obesity,[41] smoking,[42] and family history[43] (level of evidence III-3, IV). A few studies have conducted serological assays comparing the prevalence of anti-*Helicobacter pylori* antibodies between BE cases and controls, reporting risk reductions of about 50% for persons with

past infection with *H. pylori*.[44,45] There is no evidence that alcohol consumption or dietary or nutritional factors influence risk.[46,47]

Recommendation. Clinical assessment of a person's future risk of BE should consider their age, sex, history of gastro-esophageal acid reflux, waist–hip ratio, or other measures of central adiposity, smoking history, and family history of EAC and/or BE (grade B).

What is the incidence of neoplasia in patients with BE? Five population-based, prospective studies with large sample sizes and complete follow up of patients with uncomplicated BE with no dysplasia have reported progression rates to high-grade dysplasia (HGD) or adenocarcinoma of 2.2–2.6/1000 person-years (py) in Northern Ireland,[48,49] 3.3/1000 py in the Netherlands,[50] 1.2/1000 py in Denmark,[51] and 3/1000 py in the United Kingdom.[52] Meta-analyses of high-quality studies derived similar estimates of progression risks.[53,54]

What are the risk factors for progression from non-dysplastic BE to HGD or adenocarcinoma? Increased rates of progression from non-dysplastic BE to HGD or adenocarcinoma have been associated with patient factors (age, sex, smoking), endoscopic appearance (greater segment length), and aneuploidy[48,55–58] (level of evidence III-2). There is observational evidence that regular users of proton pump inhibitors (PPIs), non-steroidal anti-inflammatory drugs, and statins may have lower rates of progression from BE to cancer[59–64] (level of evidence: II, III-2, III-3).

Recommendation. Clinical assessment of future risk of HGD or adenocarcinoma in the setting of non-dysplastic BE should consider age, sex, smoking history, and endoscopic findings (grade C).

For which populations is screening for BE cost-effective? In line with accepted epidemiologic practice, these guidelines reserve "screening" to describe the process of identifying new cases of disease in an unselected population, whereas "surveillance" describes the systematic follow up of patients with known disease at periodic intervals as part of an early detection strategy to prevent progression to cancer.

There is no evidence to support population screening for BE. However, health economic studies generally suggest that one-off screening of 50-year-old men with gastro-esophageal reflux disease might be cost-effective. Both the cytosponge[65] and ultra-thin endoscopy[66] may be more cost-effective compared with standard endoscopic screening. General population screening, even if conducted coincident with colonoscopy screening, is not cost-effective.

What is appropriate medical systemic therapy for symptoms associated with BE? Medical systemic therapy for patients with BE aims to control symptoms and reduce the risk of complications. Uncomplicated BE is not a cause of symptoms (indeed patients with BE may have reduced sensitivity to esophageal acidification); rather these are due to the symptoms of gastro-esophageal reflux.[67] Acid suppression with PPI is the most effective systemic therapy for reflux symptoms in patients with BE and will control symptoms in most patients with a durable effect over years (level of evidence II, IV)[68–78] Higher than standard doses of PPI may be required to control symptoms in a proportion of patients (level of evidence IV).[79–81]

Recommendation. Symptomatic patients with BE should be treated with PPI therapy, with the dose titrated to control symptoms (grade C).

Are there any medical or surgical interventions that cause regression of BE? Regression of BE is defined by a reduction in the length or area of metaplastic columnar epithelium; however, the significance of regression in BE is unclear. There are insufficient data to indicate that regression leads to reduced incidence of EAC. The degree of Barrett's regression appears largest among patients undergoing anti-reflux surgery although a randomized trial comparing surgical and medical therapy found no significant differences.[76]

Combined analysis of randomized trials has not demonstrated BE regression with medical therapy[82] (level of evidence I).

Recommendation. There is insufficient evidence to recommend the use of acid suppressive therapy for the regression of BE (grade B).

There is insufficient evidence to recommend anti-reflux surgery for the regression of BE (grade C).

Practice point. Acid suppressive therapy and anti-reflux surgery can be used to control symptoms and heal reflux esophagitis in patients with BE. There is insufficient evidence to recommend high-dose (twice daily) acid suppressive therapy when symptom control or mucosal healing is achieved with standard dosing.

Is there a role for ablative therapy to treat BE? Various endoscopic techniques have been investigated for eradicating BE epithelium, including those that deliver focal ablation (argon plasma coagulation [APC], laser heater probe, and endoscopic mucosal resection [EMR]) and those that ablate broad fields (photodynamic therapy [PDT] and radiofrequency ablation [RFA]).

APC is a widely available monopolar electrocautery method. Randomized trials show that medically treated patients and patients with prior fundoplication can be cleared of Barrett's mucosa whereas control patients do not show significant regression.[83–85]

PDT involves administration of a photosensitizer drug (typically oral aminolevulinic acid, or IV photofrin) and subsequent exposure of the Barrett's mucosa to a laser light. Because of potentially severe skin sensitivity, the subject must remain in a darkened environment, restricting use of this technology to cooler climate countries.

RFA involves placement of a balloon catheter in the esophagus, through which radiofrequency energy is delivered allowing treatment of a 3-cm circumferential segment of the esophagus.

Side effects include chest pain, dysphagia, and stricture formation. Rare complications such as bleeding and perforation have been noted. Randomized sham-controlled studies have shown high levels of eradication of both non-dysplastic (> 90%) and dysplastic (> 90%) Barrett's mucosa.[82] Long-term follow up studies show the response is durable with the majority of patients (> 85%) maintaining complete eradication at 5 years.

Recommendation. Long-term outcome studies do not yet support ablation in patients without dysplasia (grade B).

Are there any treatments that prevent progression of BE to cancer? There is limited evidence to support preventive strategies. The choice of anti-reflux therapy (i.e. PPIs *vs* anti-reflux surgery) has not been shown to influence progression to cancer. There is interest in the use of COX inhibitors, but to date only small trials have been conducted with no clear evidence of benefit. A large randomized controlled trial is being conducted to evaluate the efficacy of aspirin to prevent the onset of cancer in patients with BE.[86] This trial is due to report in 2019.

Ablation therapies have shown benefit in randomized trials, but only in those who have already developed dysplasia. In these individuals, the risk of cancer progression appears to be reduced by approximately 50% by both PDT[87] and RFA,[88–90] but cancer risk is not eliminated. The only randomized trial[91] to evaluate ablation (APC) in non-dysplastic BE failed to show benefit for ablation.

Recommendation. Ablation of BE should remain limited to individuals with HGD in BE who are at imminent risk of developing EAC (grade B).

Practice points. The treatment of gastro-esophageal reflux with either PPIs or anti-reflux surgery has not been shown to influence progression to EAC.

There is currently no high-quality evidence supporting the use of COX inhibitors for prevention of EAC.

How frequently should patients with BE undergo endoscopy? The aim of surveillance is to detect dysplasia and early cancer for early treatment. Endoscopic surveillance in patients with BE is the current standard of practice,[8,9] although there is no evidence from randomized controlled trials for its effectiveness. There is, however, indirect evidence based on earlier stage and improved survival in EAC patients detected at surveillance, although these retrospective studies are subject to potential lead and length time bias.[92,93]

Both the British Society of Gastroenterology (BSG) and American Gastroenterological Association (AGA) have published guidelines for endoscopic surveillance of BE.[8,9] The guidelines differ in the criteria for the diagnosis of BE with both requiring a CLE but the AGA also requiring intestinal metaplasia to be present in biopsies from the CLE. This Australian guideline uses the AGA criteria for a diagnosis of BE. British and American guidelines also use the grade of dysplasia found at endoscopy to determine the timing of the subsequent surveillance endoscopy. These recommendations are based on the evidence of an increased risk of EAC with increasing degrees of dysplasia. In those with no dysplasia, the BSG guidelines also take into account the absence of intestinal metaplasia and short-segment (< 3 cm) length, both of which appear to be associated with a decreased risk of malignant progression. Both guidelines recommend biopsies of any visible lesion or mucosal irregularity and quadrantic biopsies. The BSG guidelines recommend quadrantic biopsies every 2 cm in all surveillance endoscopies. The AGA guidelines recommend Seattle protocol biopsies with quadrantic biopsies every 2 cm unless there is suspected or known dysplasia where every 1 cm is recommended. These biopsy protocols have been shown to increase the detection of advanced (high grade and early adenocarcinoma) lesions.[94,95] However, there is low adherence to the protocols[96] resulting in lower detection rates of dysplasia.[97]

The recommendations of the Australian working group for frequency of surveillance are shown in Table 3 and Figure 1. The diagnosis of BE requires intestinal metaplasia in biopsies from the CLE. Recommendations for CLE without intestinal metaplasia are discussed below.

Uncertainty regarding risk of low-grade dysplasia (LGD) progression. The optimum management of patients diagnosed with LGD is uncertain. There is considerable debate about the risks of progression to HGD or cancer in this group. Population-based studies report cancer progression rates of ~0.5% p.a.[51] In contrast, studies undertaken in academic centers in which diagnoses of LGD are made only after review by expert gastrointestinal pathologists report progression rates up to 13% p.a.[98] Importantly, in those studies, about 85% of patients diagnosed originally with LGD were down-staged to non-dysplastic BE upon expert review. Among down-staged patients, the progression rate was ~0.5% p.a.

Endoscopic surveillance in patients with CLE without intestinal metaplasia. In patients with no intestinal metaplasia or dysplasia detected in biopsies from long-segment (≥ 3 cm) CLE, endoscopic surveillance as per the protocol for long-segment BE is recommended (i.e. every 2–3 years). If there is 1 to < 3 cm of CLE without intestinal metaplasia or dysplasia, a repeat endoscopy in 3–5 years is suggested with consideration for discharge from surveillance if the repeat endoscopy with Seattle protocol biopsies again shows no intestinal metaplasia or dysplasia. In patients with CLE less than 1 cm without intestinal metaplasia or dysplasia on biopsies from the CLE, no endoscopic surveillance is suggested. If dysplasia is found in any biopsies from a CLE without intestinal metaplasia, then recommendations are as per the protocols for BE with dysplasia.

Practice points. In the absence of randomized trial evidence, the frequency of surveillance endoscopy in BE can be guided by current practice guidelines.

It is advisable to undertake endoscopic surveillance in suitable patients with BE. The frequency of surveillance is based on the presence or absence of dysplasia on previous Seattle protocol biopsies and length of BE.

A diagnosis of dysplasia (indefinite, low, and high grade) should be confirmed by a second pathologist, ideally an expert gastrointestinal pathologist.

Esophageal biopsies should be taken according to the Seattle protocol.

Table 3 Recommended frequency of endoscopic surveillance of patients with Barrett's esophagus

No dysplasia[†] on endoscopic assessment and Seattle protocol biopsy[‡]	
Short (< 3 cm) segment	Repeat endoscopy in 3–5 years
Long (≥ 3 cm) segment	Repeat endoscopy in 2–3 years

[†]If there has been previous low-grade dysplasia, see *low-grade dysplasia* protocol.

[‡]*Seattle protocol*—biopsy of any mucosal irregularity and quadrantic biopsies every 2 cm unless known or suspected dysplasia then quadrantic biopsies every 1 cm.

Indefinite for dysplasia on biopsy

The changes of *indefinite for dysplasia* on biopsy should be confirmed by a second pathologist, ideally an expert gastrointestinal pathologist. If *indefinite for dysplasia* is confirmed, then the following endoscopic surveillance is recommended:

1. Repeat endoscopy in *6 months* with Seattle protocol biopsies for suspected dysplasia (biopsy of any mucosal irregularity and quadrantic biopsies every 1 cm) on maximal acid suppression.
2. If repeat shows *no dysplasia*, then follow as per non-dysplastic protocol.
3. If repeat shows *low-grade* or *high-grade dysplasia* or *adenocarcinoma*, then follow protocols for these respective conditions.
4. If repeat again shows confirmed *indefinite for dysplasia*, then repeat endoscopy in 6 months with Seattle protocol biopsies for suspected dysplasia.

Low-grade dysplasia on biopsy

The changes of *low-grade dysplasia* on biopsy should be confirmed by a second pathologist, ideally an expert gastrointestinal pathologist. If *low-grade dysplasia* is confirmed, then the following endoscopic surveillance is recommended (or refer to an expert center for assessment):

1. Repeat endoscopy every *6 months* with Seattle protocol biopsies for dysplasia (biopsy of any mucosal irregularity and quadrantic biopsies every 1 cm).
2. If two consecutive 6 monthly endoscopies with Seattle dysplasia biopsy protocol show *no dysplasia*, then consider reverting to a less frequent follow up schedule.

High-grade dysplasia or adenocarcinoma on biopsy

Referral to a center that has integrated expertise in endoscopy, imaging, surgery, and histopathology.

Is surveillance cost-effective for follow up of patients with BE? A recent systematic review[99] of seven studies[100–106] found inconsistent assessments of the value of surveillance, ranging from being cost-effective to highly cost-ineffective. Hence, surveillance of all patients with non-dysplastic BE may not be cost-effective, but this may change with identification of patients at high risk of progression to EAC.

Are there groups of patients with non-dysplastic BE that require more frequent surveillance? Surveillance protocols for patients with BE are based on observational studies.[54,107] However, groups of patients may be identified with high rates of progression, and thus who may benefit from more frequent surveillance. Such groups include patients with longer segments of BE (≥ 3 cm) (level of evidence III-2),[53,54,56,107–110] as well as older patients, males, and smokers (level of evidence II, III-2).[48,55,57,111,112]

Recommendation. Patients with BE length equal to or greater than 3 cm may have intensive surveillance, possibly every 2–3 years following the Seattle protocol (grade C).

Are there groups of patients with BE that can be discharged from surveillance? There is limited high-quality evidence to address this question with certainty, although studies are in progress which may yield risk reducing modifiers (II, III-2, III-3).

Recommendation. For patients with < 1 cm of CLE that do not have evidence of intestinal metaplasia or dysplasia on Seattle protocol biopsy of the segment, endoscopic surveillance is not recommended (grade C).

Practice point. Patients with evidence of "regression" of BE (i.e. reduced CLE length or absence of intestinal metaplasia) can still continue surveillance.

Patients with significant comorbidities, or those unable to tolerate procedural intervention for dysplasia/EAC, may be considered for discharge from surveillance.

Guidelines for BE with dysplasia or early cancer

What are the endoscopic features of neoplasia (dysplasia and early cancer) within a BE segment? Because random sampling of quadrantic biopsies every 2 cm suffers from sampling error and, at times, limited adherence,[97,113] newer modalities have been proposed including chromoendoscopy, electronic image enhancement technologies, and high magnification platforms. There is limited information whether these methods can ultimately change patient management. Presently, high-resolution white light endoscopy (HR-WLE) remains the gold standard in evaluating patients with BE although the newer modalities may be used in addition to HR-WLE to

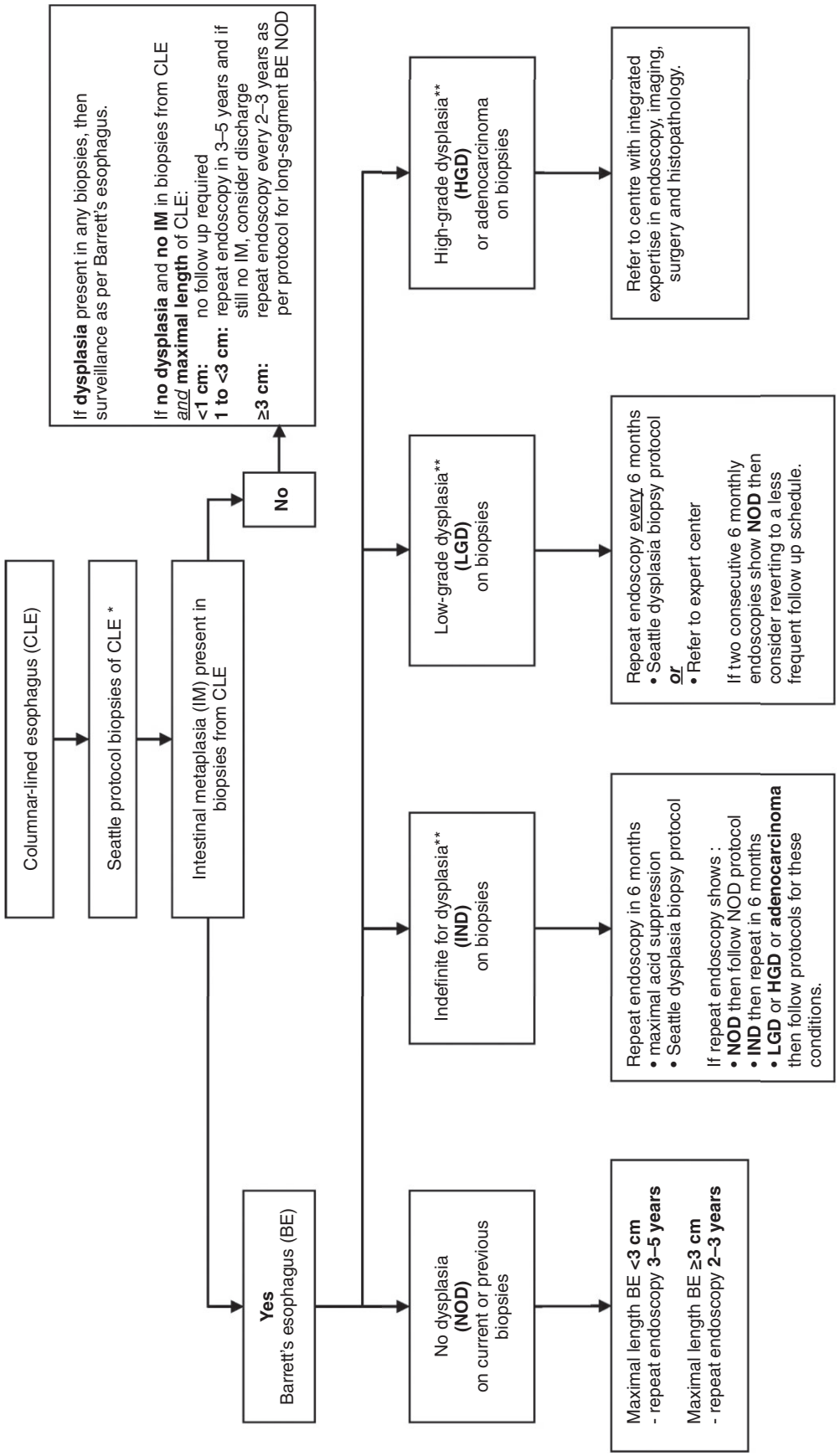


Figure 1 Decision tree for the management of Barrett's esophagus. *Seattle protocol—biopsy of any mucosal irregularity and quadrantic biopsies every 2 cm unless known or suspected dysplasia then quadrantic biopsies every 1 cm. **Dysplasia (indefinite, low, and high grade) should be confirmed by a second pathologist, ideally an expert gastrointestinal pathologist.

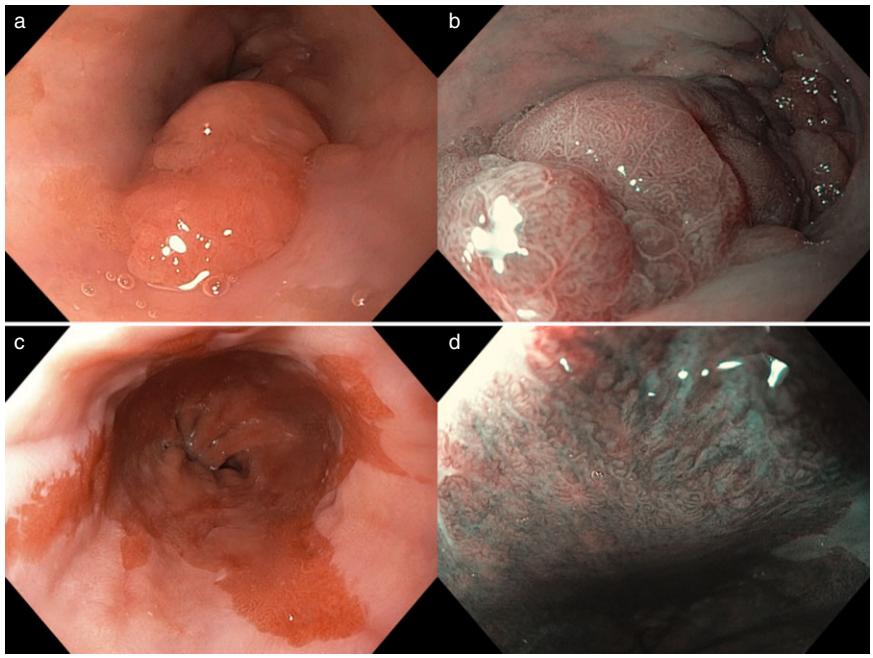


Figure 2 (a) C0M3 Barrett's esophagus containing a 2 × 1 cm (Paris 0–Is) lesion at 6 o'clock in white light and in (b) as seen with narrow-band imaging. (c) Flat C2M4 Barrett's esophagus. (d) Closer examination using narrow-band imaging reveals a focal area with irregular capillary and mucosal pattern at 12 o'clock.

improve characterization of lesions.[114] Thus, it is important to understand the gross morphological features of dysplasia and early cancer and if available, apply some of the more advanced imaging methods.

Given the inconspicuous nature of dysplasia in BE,[115] meticulous inspection and attention to subtle endoscopic anomalies using the best available imaging equipment and endoscopes are warranted. Debris and mucus should be washed off. If there is extensive peristalsis, antispasmodic agents can be used. There is some evidence that cancer preferentially occurs in the distal Barrett's segment[116] and in the 2–5 o'clock position in patients with shorter segments of BE (< 5 cm).[117]

All ulcers in BE should be monitored closely for carcinoma. Biopsies should always be taken in depressed regions and if negative, repeated after a course of PPI therapy. Visible lumps or nodules consisting of HGD suggest a more advanced lesion where more sinister pathology may be present. Suspicious lesions visualized on "white light overview" can be interrogated further with any of the enhanced imaging techniques described earlier. It is not yet clear, however, whether these modalities can replace biopsies (Fig. 2).

What is the histological definition and grading of dysplasia in patients with BE? Dysplasia is an unequivocal neoplastic transformation of the epithelial cells that is confined within the basement membrane of the metaplastic glandular tissue within which it arises. Histological features that characterize dysplasia are best identified on standard H&E-stained sections and comprise cytological changes and/or architectural changes.[118,119]

Cytological features involve nuclear changes (such as increase in size, irregular shape, increased nuclear:cytoplasmic ratio, nuclear crowding, hyperchromasia, and the presence of nucleoli) and cytoplasmic changes such as mucin depletion. Dysplastic cells exhibit increased mitotic activity, including atypical forms and surface mitoses. There is typically failure of cellular maturation toward the surface of the mucosa, although this is not always the case.[120] Goblet cell numbers are reduced and dysplastic cells may lose their normal vertical polarity.

Architectural features are irregular gland outline, variability in glandular size, gland crowding with "back-to-back" pattern, and villiform surface contour. None of these cytological or architectural features are sufficient to diagnose dysplasia in isolation. Ancillary tests (e.g. p53, AMACR and Ki67 stains) have been advocated to aid the diagnosis of dysplasia; however at present, conventional H&E examination remains the gold standard.

Grading of BE dysplasia is best performed on the H&E stain. Pathologists should report BE biopsies as fitting into one of four categories.[118,119,121–123] The rationale for this tiered approach is to stratify patients into categories of increasing risk for development of or concurrent presence of EAC. Many papers have shown an increasing risk ranging from small (negative for dysplasia) to significant (HGD).[124]

1 Negative for dysplasia
2 Indefinite for dysplasia—when the pathologist believes that the biopsy is displaying some features of true dysplasia but is unable to exclude a non-neoplastic process as the cause of the abnormality. In general, the consideration is whether the histological features are sufficient to diagnose LGD. However, in some situations the pathologist is concerned that the features may represent HGD. The concept of indefinite for HGD/adenocarcinoma has not been studied specifically; however, pathologists recognize a subgroup of indefinite for dysplasia where the cytological and/or architectural abnormality is marked but a confident diagnosis of HGD cannot be made. In some of these situations, the concern is that invasive adenocarcinoma may exist.
3 LGD—displays mild-to-moderate cytological atypia and, at

most, mild disturbance of gland architecture. The neoplastic epithelial cells are crowded, elongated, and hyperchromatic. The cells generally retain their vertical polarity.

4 HGD—typically displays both architectural abnormality and severe cytological atypia. Aberrant architectural features include glandular crowding, branching or budding glands, villiform, cribriform, micropapillary, or cystically dilated crypt patterns. Cytological features include complete loss of cell polarity, rounded enlarged nuclei with irregular-thickened nuclear membranes, and conspicuous nucleoli. Typical and atypical mitotic figures are readily identified at all levels within the glands, as well as on the luminal surface.

Grading of dysplasia is subject to significant interobserver variability,[125–127] especially LGD. Interobserver agreement among general histopathologists ranges from kappa values of 0.14 to 0.32. Specialist gastrointestinal histopathologists have better agreement (kappa 0.48–0.69).[128] When a diagnosis of LGD made by a general histopathologist is reviewed by an expert panel, the diagnosis is most often down-graded to "negative for dysplasia."

These data support the notion that all cases of BE diagnosed as dysplasia (indefinite, low, or high grade) should be reviewed by at least one expert GI pathologist.

What are the histological features of early adenocarcinoma of the esophagus? Early adenocarcinoma refers to invasion into mucosa or superficial submucosa, but not deeper (T1 in the current TNM system). Adenocarcinoma exists when there is invasion beyond the basement membrane of the epithelium. The histological features identifying that invasion has occurred include:[129,130]

1 Single neoplastic cells or small clusters of neoplastic cells in the lamina propria.
2 Complex architectural patterns characterized by solid growth patterns, tight cribriform growth pattern, glands with acute angulation in at least one part of their outline, and a pattern of anastomosing fusion of small glands.
3 Neoplastic cells invading overlying squamous epithelium.
4 Desmoplastic stromal reaction.

Significant interobserver variability exists between pathologists in the separation of HGD from early invasive adenocarcinoma in biopsy specimens.[131] Recent studies have identified a variety of histological patterns that predict invasive adenocarcinoma including solid or cribriform growth patterns, ulceration of dysplastic epithelium, abundant neutrophils within dysplastic epithelium, dilated neoplastic glands containing necrotic debris, and dysplastic glandular epithelium being incorporated into squamous epithelium. The risk of adenocarcinoma is increased with number of features present.[132]

The histological report of EMRs should include data that are important for clinical management, particularly the identification of patients who should be considered for esophagectomy. These are discussed in greater detail in the guidelines for reporting esophageal and gastro-esophageal carcinomas provided by the Royal College of Pathologists of Australasia.[133]

What are the best modalities for accurately staging early EAC? Early EACs are those defined as intramucosal adenocarcinoma (T1a) or superficial submucosal adenocarcinoma (T1b).[114] A more comprehensive subclassification of early esophageal cancers has been proposed with mucosal disease and submucosal disease divided into three categories, respectively (m1-3/4 and sm1-3) based on depth of invasion.

Options for staging of early EAC include:

1 Endoscopic biopsy
2 Endoscopic resection (ER) (also known as EMR)
3 Endoscopic ultrasound (EUS) with or without fine-needle aspirate (FNA)
4 Positron emission tomography-computerized tomography (PET-CT), once the diagnosis of cancer has been confirmed

Endoscopic biopsy is useful but is subject to sampling error. ER is superior to biopsy and results in a change in diagnosis in up to 50% of patients with dysplasia or adenocarcinoma (level of evidence IV).[134–137] Moreover, ER allows improved pathological staging of HGD and T1m and T1sm adenocarcinoma as compared with biopsy and EUS (level of evidence IV)[136,138] (Fig. 3). Rates of adverse events following ER, such as perforation, bleeding, and stricturing, are low when performed at expert centers (level of evidence IV).[138–140] EUS is not accurate for determining the stage of early EAC, especially distinguishing T1m from T1sm tumors. It is useful for differentiating T1 and >T1 stages (level of evidence IV).[141,142] EUS and EUS guided FNA (EUS-FNA) are superior to CT for locoregional lymph node staging (level of evidence IV).[143,144]

Recommendations. ER is the most accurate staging modality for early EAC for suitable lesions and where appropriate expertise is available (grade D).

EUS can be used prior to ER when deeper invasion is considered likely, particularly for lesions with ulcerated or depressed morphology (grade D).

FDG-PET or PET/CT is not routinely indicated in staging early EAC. It is best used for the staging of distant metastases or in cases of suspected more advanced local disease (grade D).

What is the appropriate management of LGD in patients with BE? Recent studies suggest that when the diagnosis of LGD is agreed on by two or more expert pathologists, the risk of progression to neoplasia is higher than previously reported (level of evidence III-2).[88,98,145] British and American guidelines recommend increased frequency of surveillance.[8,9] Endoscopic ablation with a range of methods is associated with lower rates of progression to cancer (level of evidence IV).[146] In particular, an RCT reported that RFA in patients with confirmed LGD have significantly lower rates of progression to cancer or HGD, although as yet there is no evidence of an overall survival benefit (level of evidence II).[88]

Recommendations. The diagnosis of LGD should be confirmed by a second pathologist, ideally an expert gastrointestinal pathologist (grade C).

In patients with confirmed LGD, it is advised to perform rigorous high-definition endoscopy or refer to an expert centre for assessment (grade C).

In patients with confirmed LGD, intensified endoscopic surveillance is required. Endoscopic ablation may be considered

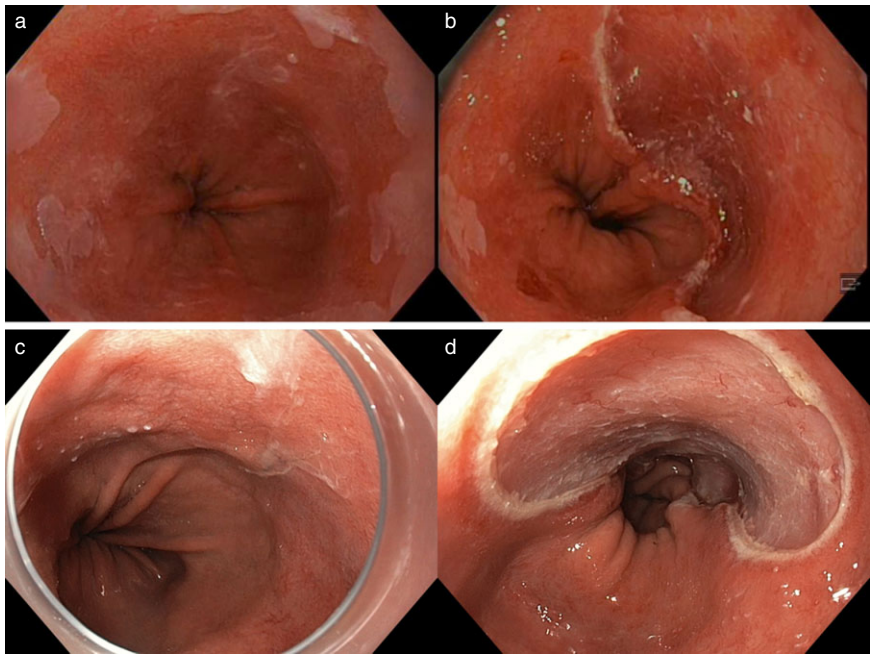


Figure 3 (a) C3M4 Barrett's esophagus. After careful inspection, a focal abnormality was noted at 2 o'clock. (b) Focal endoscopic mucosal resection was performed for staging confirming high-grade dysplasia. (c) C7M8 Barrett's esophagus. Using a distal attachment cap for improved visualization, nodular lesion with slight depression (Paris 0–IIa+IIc) noted at 12–2 o'clock. (d) This area is completely excised by endoscopic mucosal resection. Histology confirmed Barrett's esophagus with high-grade dysplasia and focal area of intramucosal adenocarcinoma (M1-T1a).

especially where LGD is definite, multifocal, and present on more than one occasion. This decision needs to be individualized based on discussion of risk and benefits with the patient (grade B).

What are the goals of treatment of HGD in patients with BE? There is no high-level evidence that directly answers this question, and so the guidelines are based on expert opinion. As HGD is prone to both over- and under-staging, the first goal of management is to confirm the diagnosis.

Once HGD has been confirmed, the goal of treatment is to prevent the progression to malignancy through the removal of dysplastic tissue. More specifically, the goals of treatment are:

1 The removal of all dysplastic tissue[114]
2 The removal of all Barrett's metaplasia if possible[114]
3 Preservation of normal swallowing/nutrition
4 Minimization of morbidity due to the eradication technique
5 Confirmation of the diagnosis of HGD (i.e. exclusion of malignancy) through examination of resected tissue (endoscopically or surgically), where possible
6 Continued follow up in patients who have had endoscopic therapy[114]

There is no management strategy that perfectly fulfils all these criteria. Current practice favors endotherapy (ER or ablation) over surveillance or esophagectomy for HGD/T1a cancer, although no randomized control trials have compared the two modalities directly. All patients should be discussed at a multidisciplinary meeting.

Practice point. The confirmation of HGD should act as a trigger for definitive treatment.

What is the best endoscopic treatment for HGD in patients with BE? ER alters histological grade or local T stage in 48% of patients and reduces esophagectomy rates by providing an effective local therapy. ER has a high success rate (94%) for complete Barrett's excision in short-segment BE (level of evidence IV).[139] RFA has been shown to completely eradicate HGD in 81% of patients at 1 year of follow up versus 19% complete eradication in patients undergoing endoscopic surveillance alone. Similar outcomes are reported following RFA at 2 and 3 years of follow up with 95% and 96% complete eradication, respectively (level of evidence II)[89,90] (Fig. 4).

Recommendations. ER should be considered for patients with intramucosal adenocarcinoma or HGD and visible/nodular lesions (grade D).

RFA should be considered for patients with HGD within flat segments of BE. RFA is not appropriate in patients with visible abnormalities; these should be treated by ER. RFA may be the preferred treatment strategy over ER for patients with long-segment BE or circumferential Barrett's due to a lower rate of stricture formation (grade B).

Practice point. It is advisable to refer patients with BE and dysplasia or early EAC to tertiary referral centers for management.

What is the best endoscopic management of early EAC? Early EAC comprises the histological tumor classification of T1a (invasion into the mucosa) and T1b (invasion into submucosa but not muscularis propria). The depth of invasion can be further stratified based on mucosal (m1–m3/m1–m4) or submucosal (sm1–sm3) involvement.[123,147] ER is the most accurate T staging modality for early EAC (level of evidence IV)[137,139]

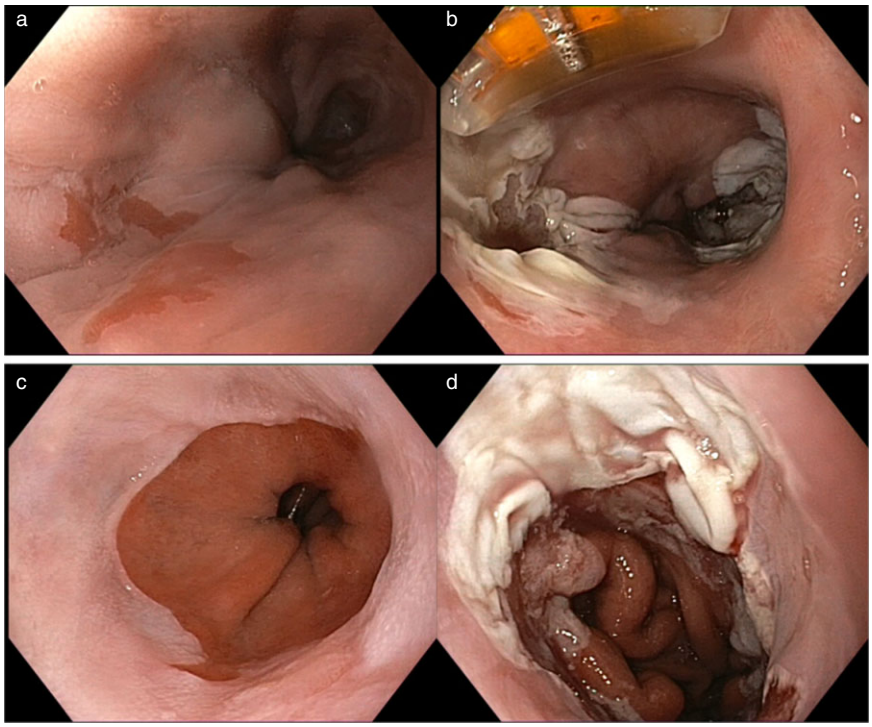


Figure 4 (a) C5M7 Barrett's esophagus with high-grade dysplasia previously treated by endoscopic mucosal resection and radiofrequency ablation—residual disease remaining at 7 o'clock proximally and 12–4 o'clock distally. (b) Focal radiofrequency ablation to sites of residual Barrett's mucosa. (c) C2M4 Barrett's esophagus previously treated by radiofrequency ablation for flat high-grade dysplasia. (d) Residual Barrett's mucosa is treated by focal radiofrequency ablation.

(Fig. 5). The risk of lymph node involvement with T1a and T1b early EAC is 1.3–2.5% and 12–31%, respectively.[148–151] Unlike locally advanced or node-involving disease, early EAC can often be cured with surgical or endoscopic approaches. Endoscopic treatment is less morbid and expensive than esophagectomy, and is organ preserving.[152] ER is effective for T1a early EAC when performed in experienced centers. Selected patients with T1b early EAC may benefit from ER if esophagectomy is not indicated (levels of evidence II, III-2, IV).[153–157]

Recommendations. All lesions and visible abnormalities should be staged by focal ER (grade D).

If ER of early EAC is planned, ER is appropriate in most cases. Ablative therapies should not be used as primary endoscopic therapy for early EAC (grade C).

Patients with T1a adenocarcinoma on endoscopic work-up should be offered ER in preference to esophagectomy (grade D). Selected patients with T1b early EAC may also be offered ER but only if esophagectomy is not indicated (grade D).

Following resection of early EAC the remaining Barrett's mucosa should be eradicated. Barrett's eradication options include complete ER, RFA, cryotherapy, and APC (grade C).

Following resection of early EAC, the patient should undergo regular and careful surveillance examinations (grade C).

Practice point. ER of early EAC should be performed in referral centers that have integrated expertise in endoscopy, imaging, surgery, and histopathology.

Careful and dedicated endoscopic interrogation of all Barrett's mucosa is advised.

After successful endoscopic treatment for BE neoplasia, how frequently should patients undergo endoscopy? There is no high-level evidence that directly answers this question, and so the guidelines are based on expert opinion. Following endoscopic treatment for BE with neoplasia, patients should be considered for three monthly surveillance endoscopies with Seattle protocol to confirm clearance of disease. Once clearance has been achieved, consider six monthly endoscopic surveillance for 1 year, then annually. Higher risk patients may require closer surveillance endoscopy after clearance of BE neoplasia is achieved (i.e. initially six monthly for a year). ER of mucosal irregularities (nodules, depressed areas) in the squamous epithelium should be considered to clarify possible recurrent or metachronous intramucosal adenocarcinoma from subsquamous glands.

Practice point. Consider three monthly surveillance endoscopy with Seattle protocol during the endoscopic treatment phase to confirm clearance of intramucosal adenocarcinoma and residual BE. Once clearance has been achieved, consider six monthly endoscopic surveillance for 1 year, then annually.

Higher risk patients may require closer surveillance endoscopy after clearance of BE neoplasia is achieved (i.e. initially three monthly for a year). ER of any nodularity in the squamous epithelium should be considered to clarify possible recurrent or metachronous cancer from subsquamous glands.

What endoscopic surveillance protocol should be followed for patients with HGD? Surveillance is generally not indicated for patients with HGD and therapeutic intervention must be considered instead.

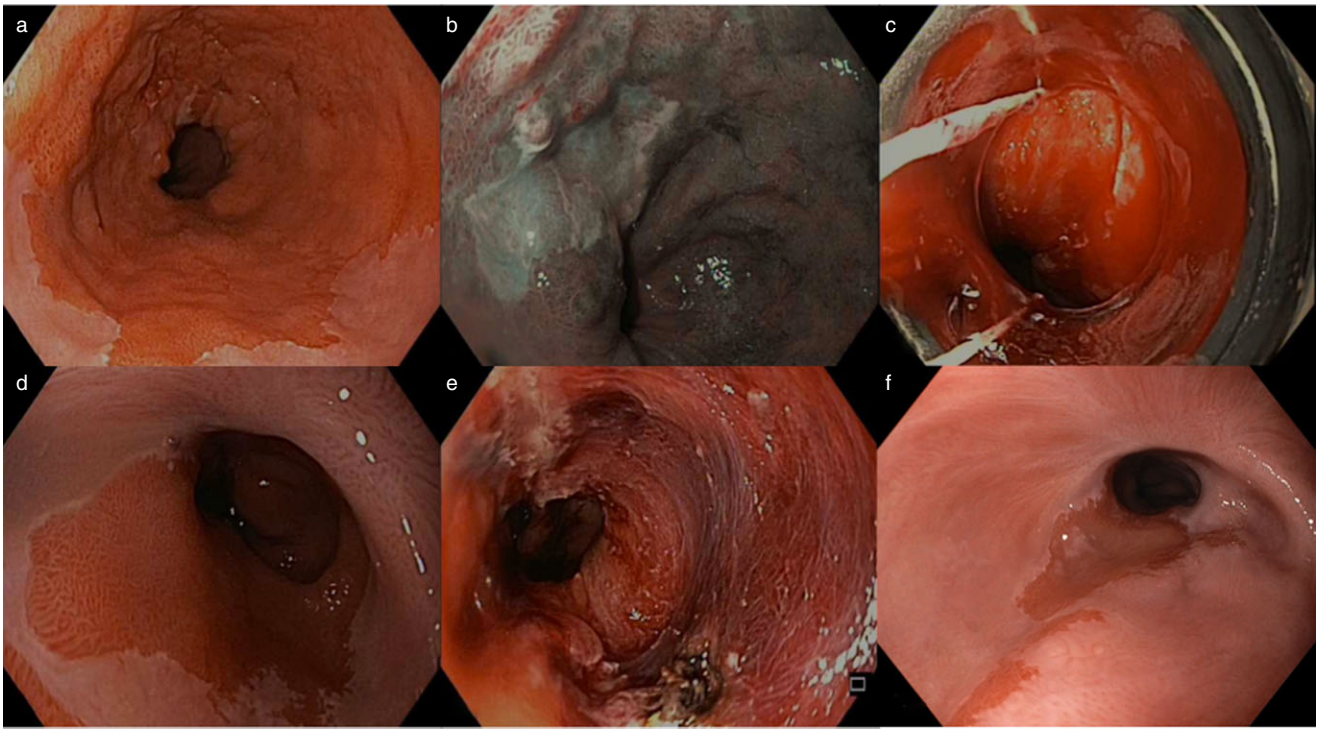


Figure 5 (a) C4M5 Barrett's esophagus with diffuse nodular mucosa between 9 and 2 o'clock. (b) Close up examination using narrow-band imaging discloses irregular mucosa with abnormal capillary and pit patterns. (c) The abnormal area is removed by multiband mucosectomy; this revealed intramucosal adenocarcinoma (T1a). (d) On progress examination at 6 weeks neosquamous epithelium is seen in the area of excision. (e) Further, stepwise complete endoscopic resection for the residual Barrett's mucosa is performed resulting in a hemi-circumferential mucosal defect. (f) Six weeks following the previous resection there is extensive neo-squamous re-epithelialization of the distal esophagus. A small residual segment of Barrett's mucosa remains. The patient had low-grade dysphagia from this stricture that was easily treated with Savary dilation.

How effective is endoscopic management compared with surgical management for HGD in patients with BE? There are no randomized controlled trials comparing surgery with endoscopic treatments for HGD. Evidence therefore comes largely from non-randomized retrospective studies. These studies report that endoscopic treatment of HGD provides similar outcomes to surgery with regard to overall survival and cancer-related mortality (level of evidence III-2).[153,158–162] In addition, the studies tend to report that compared with surgery, endoscopic treatments result in less morbidity but higher rates of local recurrence (level of evidence III-2).[153,158–162]

Recommendation. Patients with HGD in BE should be managed in centers with high-volume experience of the condition. The treatment and follow up should occur in those specialist centers (grade C).

Practice points. Patients with HGD in BE can be discussed at a multidisciplinary team meeting at a specialist centre.

Endoscopic treatment will be the first-line treatment option for the majority of patients with HGD in BE. There will be a group of patients for whom endoscopic treatment is not appropriate or successful and they will be best treated with surgery in a specialist centre.

Acknowledgments

The guidelines development process was supported by the Cancer Council Australia (CCA). We thank the CCA guidelines team for their work in supporting this process.

References

1 Cancer Council Australia Barrett's Oesophagus Guidelines Working Party. *Clinical Practice Guidelines for the Diagnosis and Management of Barrett's Oesophagus and Early Oesophageal Adenocarcinoma*. Cited 17 February 2015. Available from URL: http://wiki.cancer.org.au/australia/Guidelines:Barrett%27s
2 Cancer Council Australia. *Development of Clinical Practice Guidelines Using Cancer Council Australia's Cancer Guidelines Wiki. Handbook for Section Authors and the Guideline Working Party*. Sydney: Cancer Council Australia, 2014.
3 Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am. J. Gastroenterol.* 2006; **101**: 1900–20.
4 Shaheen NJ, Richter JE. Barrett's oesophagus. *Lancet* 2009; **373**: 850–61.
5 Ishimura N, Amano Y, Appelman HD *et al*. Barrett's esophagus: endoscopic diagnosis. *Ann. N. Y. Acad. Sci.* 2011; **1232**: 53–75.

6 Sharma P, Dent J, Armstrong D *et al*. The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. *Gastroenterology* 2006; **131**: 1392–9.
7 Paull A, Trier JS, Dalton MD, Camp RC, Loeb P, Goyal RK. The histologic spectrum of Barrett's esophagus. *N. Engl. J. Med.* 1976; **295**: 476–80.
8 Fitzgerald RC, di Pietro M, Ragunath K *et al*. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* 2014; **63**: 7–42.
9 Spechler SJ, Sharma P, Souza RF, Inadomi JM, Shaheen NJ. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* 2011; **140**: 1084–91.
10 Spechler SJ, Sharma P, Souza RF, Inadomi JM, Shaheen NJ. American Gastroenterological Association technical review on the management of Barrett's esophagus. *Gastroenterology* 2011; **140**: e18–52.
11 Wang KK, Sampliner RE. Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus. *Am. J. Gastroenterol.* 2008; **103**: 788–97.
12 Jung KW, Talley NJ, Romero Y *et al*. Epidemiology and natural history of intestinal metaplasia of the gastroesophageal junction and Barrett's esophagus: a population-based study. *Am. J. Gastroenterol.* 2011; **106**: 1447–55.
13 Sharma P, Weston AP, Morales T, Topalovski M, Mayo MS, Sampliner RE. Relative risk of dysplasia for patients with intestinal metaplasia in the distal oesophagus and in the gastric cardia. *Gut* 2000; **46**: 9–13.
14 Endlicher E, Rummele P, Beer S *et al*. Barrett's esophagus: a discrepancy between macroscopic and histological diagnosis. *Endoscopy* 2005; **37**: 1131–5.
15 Ferguson DD, DeVault KR, Krishna M, Loeb DS, Wolfsen HC, Wallace MB. Enhanced magnification-directed biopsies do not increase the detection of intestinal metaplasia in patients with GERD. *Am. J. Gastroenterol.* 2006; **101**: 1611–16.
16 Ngamruengphong S, Sharma VK, Das A. Diagnostic yield of methylene blue chromoendoscopy for detecting specialized intestinal metaplasia and dysplasia in Barrett's esophagus: a meta-analysis. *Gastrointest. Endosc.* 2009; **69**: 1021–8.
17 Admad NZ, Ahmed A. A meta-analysis of randomized controlled trials comparing methylene blue-directed biopsies with random biopsies in the surveillance of Barrett's esophagus. *Esophagus* 2010; **7**: 207–13.
18 Harrison R, Perry I, Haddadin W *et al*. Detection of intestinal metaplasia in Barrett's esophagus: an observational comparator study suggests the need for a minimum of eight biopsies. *Am. J. Gastroenterol.* 2007; **102**: 1154–61.
19 Gonzalez S, Yu WM, Smith MS *et al*. Randomized comparison of 3 different-sized biopsy forceps for quality of sampling in Barrett's esophagus. *Gastrointest. Endosc.* 2010; **72**: 935–40.
20 Garcia RT, Cello JP, Nguyen MH *et al*. Unsedated ultrathin EGD is well accepted when compared with conventional sedated EGD: a multicenter randomized trial. *Gastroenterology* 2003; **125**: 1606–12.
21 Jobe BA, Hunter JG, Chang EY *et al*. Office-based unsedated small-caliber endoscopy is equivalent to conventional sedated endoscopy in screening and surveillance for Barrett's esophagus: a randomized and blinded comparison. *Am. J. Gastroenterol.* 2006; **101**: 2693–703.
22 Shariff MK, Bird-Lieberman EL, O'Donovan M *et al*. Randomized crossover study comparing efficacy of transnasal endoscopy with that of standard endoscopy to detect Barrett's esophagus. *Gastrointest. Endosc.* 2012; **75**: 954–61.
23 El-Zimaity HM, Graham DY. Cytokeratin subsets for distinguishing Barrett's esophagus from intestinal metaplasia in the cardia using endoscopic biopsy specimens. *Am. J. Gastroenterol.* 2001; **96**: 1378–82.
24 Kurtkaya-Yapicier O, Gencosmanoglu R, Avsar E, Bakirci N, Tozun N, Sav A. The utility of cytokeratins 7 and 20 (CK7/20) immunohistochemistry in the distinction of short-segment Barrett esophagus from gastric intestinal metaplasia: is it reliable? *BMC Clin. Pathol.* 2003; **3**: 5.
25 Mohammed IA, Streutker CJ, Riddell RH. Utilization of cytokeratins 7 and 20 does not differentiate between Barrett's esophagus and gastric cardiac intestinal metaplasia. *Mod. Pathol.* 2002; **15**: 611–16.
26 Ormsby AH, Vaezi MF, Richter JE *et al*. Cytokeratin immunoreactivity patterns in the diagnosis of short-segment Barrett's esophagus. *Gastroenterology* 2000; **119**: 683–90.
27 Schilling D, Spiethoff A, Rosenbaum A *et al*. Does Cytokeratin7/20 immunoreactivity help to distinguish Barrett's esophagus from gastric intestinal metaplasia? Results of a prospective study of 75 patients. *Pathol. Res. Pract.* 2005; **200**: 801–5.
28 White NM, Gabril M, Ejeckam G *et al*. Barrett's esophagus and cardiac intestinal metaplasia: two conditions within the same spectrum. *Can. J. Gastroenterol.* 2008; **22**: 369–75.
29 Yim HJ, Lee SW, Choung RS *et al*. Is cytokeratin immunoreactivity useful in the diagnosis of short-segment Barrett's oesophagus in Korea? *Eur. J. Gastroenterol. Hepatol.* 2005; **17**: 611–16.
30 Glickman JN, Shahsafaei A, Odze RD. Mucin core peptide expression can help differentiate Barrett's esophagus from intestinal metaplasia of the stomach. *Am. J. Surg. Pathol.* 2003; **27**: 1357–65.
31 McIntire MG, Soucy G, Vaughan TL, Shahsafaei A, Odze RD. MUC2 is a highly specific marker of goblet cell metaplasia in the distal esophagus and gastroesophageal junction. *Am. J. Surg. Pathol.* 2011; **35**: 1007–13.
32 Groome M, Lindsay J, Ross PE, Cotton JP, Hupp TR, Dillon JF. Use of oesophageal stress response proteins as potential biomarkers in the screening for Barrett's oesophagus. *Eur. J. Gastroenterol. Hepatol.* 2008; **20**: 961–5.
33 Sipponen P, Vauhkonen M, Helske T, Kaariainen I, Harkonen M. Low circulating levels of gastrin-17 in patients with Barrett's esophagus. *World J. Gastroenterol.* 2005; **11**: 5988–92.
34 Kadri SR, Lao-Sirieix P, O'Donovan M *et al*. Acceptability and accuracy of a non-endoscopic screening test for Barrett's oesophagus in primary care: cohort study. *BMJ* 2010; **341**: c4372.
35 Lao-Sirieix P, Boussioutas A, Kadri SR *et al*. Non-endoscopic screening biomarkers for Barrett's oesophagus: from microarray analysis to the clinic. *Gut* 2009; **58**: 1451–9.
36 Nandurkar S, Talley NJ, Martin CJ, Ng TH, Adams S. Short segment Barrett's oesophagus: prevalence, diagnosis and associations. *Gut* 1997; **40**: 710–15.
37 Kendall BJ, Whiteman DC. Temporal changes in the endoscopic frequency of new cases of Barrett's esophagus in an Australian health region. *Am. J. Gastroenterol.* 2006; **101**: 1178–82.
38 Corley DA, Kubo A, Levin TR *et al*. Race, ethnicity, sex and temporal differences in Barrett's oesophagus diagnosis: a large community-based study, 1994-2006. *Gut* 2009; **58**: 182–8.
39 Cook MB, Wild CP, Forman D. A systematic review and meta-analysis of the sex ratio for Barrett's esophagus, erosive reflux disease, and nonerosive reflux disease. *Am. J. Epidemiol.* 2005; **162**: 1050–61.
40 Taylor JB, Rubenstein JH. Meta-analyses of the effect of symptoms of gastroesophageal reflux on the risk of Barrett's esophagus. *Am. J. Gastroenterol.* 2010; **105**: 1730–7.

41 Singh S, Sharma AN, Murad MH *et al*. Central adiposity is associated with increased risk of esophageal inflammation, metaplasia, and adenocarcinoma: a systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* 2013; **11**: 1399–412.
42 Cook MB, Shaheen NJ, Anderson LA *et al*. Cigarette smoking increases risk of Barrett's esophagus: an analysis of the Barrett's and Esophageal Adenocarcinoma Consortium. *Gastroenterology* 2012; **142**: 744–53.
43 Chak A, Lee T, Kinnard MF *et al*. Familial aggregation of Barrett's oesophagus, oesophageal adenocarcinoma, and oesophagogastric junctional adenocarcinoma in Caucasian adults. *Gut* 2002; **51**: 323–8.
44 Corley DA, Kubo A, Levin TR *et al*. *Helicobacter pylori* infection and the risk of Barrett's oesophagus: a community-based study. *Gut* 2008; **57**: 727–33.
45 Thrift AP, Pandeya N, Smith KJ *et al*. *Helicobacter pylori* infection and the risks of Barrett's oesophagus: a population-based case-control study. *Int. J. Cancer* 2012; **130**: 2407–16.
46 Anderson LA, Cantwell MM, Watson RG *et al*. The association between alcohol and reflux esophagitis, Barrett's esophagus, and esophageal adenocarcinoma. *Gastroenterology* 2009; **136**: 799–805.
47 Thrift AP, Pandeya N, Smith KJ *et al*. Lifetime alcohol consumption and risk of Barrett's esophagus. *Am. J. Gastroenterol.* 2011; **106**: 1220–30.
48 Bhat S, Coleman HG, Yousef F *et al*. Risk of malignant progression in Barrett's esophagus patients: results from a large population-based study. *J. Natl. Cancer Inst.* 2011; **103**: 1049–57.
49 Murray L, Watson P, Johnston B, Sloan J, Mainie IM, Gavin A. Risk of adenocarcinoma in Barrett's oesophagus: population based study. *BMJ* 2003; **327**: 534–5.
50 Schouten LJ, Steevens J, Huysentruyt CJ *et al*. Total cancer incidence and overall mortality are not increased among patients with Barrett's esophagus. *Clin. Gastroenterol. Hepatol.* 2011; **9**: 754–61.
51 Hvid-Jensen F, Pedersen L, Drewes AM, Sorensen HT, Funch-Jensen P. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N. Engl. J. Med.* 2011; **365**: 1375–83.
52 Alexandropoulou K, van Vlymen J, Reid F, Poullis A, Kang JY. Temporal trends of Barrett's oesophagus and gastro-oesophageal reflux and related oesophageal cancer over a 10-year period in England and Wales and associated proton pump inhibitor and H2RA prescriptions: a GPRD study. *Eur. J. Gastroenterol. Hepatol.* 2013; **25**: 15–21.
53 Thomas T, Abrams KR, De Caestecker JS, Robinson RJ. Meta analysis: cancer risk in Barrett's oesophagus. *Aliment. Pharmacol. Ther.* 2007; **26**: 1465–77.
54 Desai TK, Krishnan K, Samala N *et al*. The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut* 2012; **61**: 970–6.
55 Coleman HG, Bhat S, Johnston BT, McManus D, Gavin AT, Murray LJ. Tobacco smoking increases the risk of high-grade dysplasia and cancer among patients with Barrett's esophagus. *Gastroenterology* 2012; **142**: 233–40.
56 Coleman HG, Bhat SK, Murray LJ *et al*. Symptoms and endoscopic features at Barrett's esophagus diagnosis: implications for neoplastic progression risk. *Am. J. Gastroenterol.* 2014; **109**: 527–34.
57 de Jonge PJ, van Blankenstein M, Looman CW, Casparie MK, Meijer GA, Kuipers EJ. Risk of malignant progression in patients with Barrett's oesophagus: a Dutch nationwide cohort study. *Gut* 2010; **59**: 1030–6.
58 Reid BJ, Levine DS, Longton G, Blount PL, Rabinovitch PS. Predictors of progression to cancer in Barrett's esophagus: baseline histology and flow cytometry identify low- and high-risk patient subsets. *Am. J. Gastroenterol.* 2000; **95**: 1669–76.
59 Kastelein F, Spaander MC, Biermann K, Steyerberg EW, Kuipers EJ, Bruno MJ. Nonsteroidal anti-inflammatory drugs and statins have chemopreventative effects in patients with Barrett's esophagus. *Gastroenterology* 2011; **141**: 2000–8.
60 Kastelein F, Spaander MC, Steyerberg EW *et al*. Proton pump inhibitors reduce the risk of neoplastic progression in patients with Barrett's esophagus. *Clin. Gastroenterol. Hepatol.* 2013; **11**: 382–8.
61 Nguyen DM, El-Serag HB, Henderson L, Stein D, Bhattacharyya A, Sampliner RE. Medication usage and the risk of neoplasia in patients with Barrett's esophagus. *Clin. Gastroenterol. Hepatol.* 2009; **7**: 1299–304.
62 Kantor ED, Onstad L, Blount PL, Reid BJ, Vaughan TL. Use of statin medications and risk of esophageal adenocarcinoma in persons with Barrett's esophagus. *Cancer Epidemiol. Biomarkers Prev.* 2012; **21**: 456–61.
63 Singh S, Garg SK, Singh PP, Iyer PG, El-Serag HB. Acid-suppressive medications and risk of oesophageal adenocarcinoma in patients with Barrett's oesophagus: a systematic review and meta-analysis. *Gut* 2014; **63**: 1229–37.
64 Singh S, Singh AG, Singh PP, Murad MH, Iyer PG. Statins are associated with reduced risk of esophageal cancer, particularly in patients with Barrett's esophagus: a systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* 2013; **11**: 620–9.
65 Benaglia T, Sharples LD, Fitzgerald RC, Lyratzopoulos G. Health benefits and cost effectiveness of endoscopic and nonendoscopic cytosponge screening for Barrett's esophagus. *Gastroenterology* 2013; **144**: 62–73.
66 Nietert PJ, Silverstein MD, Mokhashi MS *et al*. Cost-effectiveness of screening a population with chronic gastroesophageal reflux. *Gastrointest. Endosc.* 2003; **57**: 311–18.
67 Johnson DA, Winters C, Spurling TJ, Chobanian SJ, Cattau EL, Jr. Esophageal acid sensitivity in Barrett's esophagus. *J. Clin. Gastroenterol.* 1987; **9**: 23–7.
68 Malesci A, Savarino V, Zentilin P *et al*. Partial regression of Barrett's esophagus by long-term therapy with high-dose omeprazole. *Gastrointest. Endosc.* 1996; **44**: 700–5.
69 Sontag SJ, Schnell TG, Chejfec G, Kurucar C, Karpf J, Levine G. Lansoprazole heals erosive reflux oesophagitis in patients with Barrett's oesophagus. *Aliment. Pharmacol. Ther.* 1997; **11**: 147–56.
70 Fass R, Sampliner RE, Malagon IB *et al*. Failure of oesophageal acid control in candidates for Barrett's oesophagus reversal on a very high dose of proton pump inhibitor. *Aliment. Pharmacol. Ther.* 2000; **14**: 597–602.
71 Ortiz A, Martinez de Haro LF, Parrilla P, Molina J, Bermejo J, Munitiz V. 24-h pH monitoring is necessary to assess acid reflux suppression in patients with Barrett's oesophagus undergoing treatment with proton pump inhibitors. *Br. J. Surg.* 1999; **86**: 1472–4.
72 Yeh RW, Gerson LB, Triadafilopoulos G. Efficacy of esomeprazole in controlling reflux symptoms, intraesophageal, and intragastric pH in patients with Barrett's esophagus. *Dis. Esophagus* 2003; **16**: 193–8.
73 Attwood SE, Lundell L, Hatlebakk JG *et al*. Medical or surgical management of GERD patients with Barrett's esophagus: the LOTUS trial 3-year experience. *J. Gastrointest. Surg.* 2008; **12**: 1646–54.
74 Frazzoni M, Savarino E, Manno M *et al*. Reflux patterns in patients with short-segment Barrett's oesophagus: a study using impedance-pH monitoring off and on proton pump inhibitor therapy. *Aliment. Pharmacol. Ther.* 2009; **30**: 508–15.
75 Watson JT, Moawad FJ, Veerappan GR *et al*. The dose of omeprazole required to achieve adequate intraesophageal acid

suppression in patients with gastroesophageal junction specialized intestinal metaplasia and Barrett's esophagus. *Dig. Dis. Sci.* 2013; **58**: 2253–60.
76 Parrilla P, Martinez de Haro LF, Ortiz A *et al*. Long-term results of a randomized prospective study comparing medical and surgical treatment of Barrett's esophagus. *Ann. Surg.* 2003; **237**: 291–8.
77 Sampliner RE. Effect of up to 3 years of high-dose lansoprazole on Barrett's esophagus. *Am. J. Gastroenterol.* 1994; **89**: 1844–8.
78 Zaninotto G, Parente P, Salvador R *et al*. Long-term follow-up of Barrett's epithelium: medical versus antireflux surgical therapy. *J. Gastrointest. Surg.* 2012; **16**: 7–14.
79 Basu KK, Bale R, West KP, de Caestecker JS. Persistent acid reflux and symptoms in patients with Barrett's oesophagus on proton-pump inhibitor therapy. *Eur. J. Gastroenterol. Hepatol.* 2002; **14**: 1187–92.
80 Frazzoni M, Manno M, De Micheli E, Savarino V. Efficacy in intra-oesophageal acid suppression may decrease after 2-year continuous treatment with proton pump inhibitors. *Dig. Liver Dis.* 2007; **39**: 415–21.
81 Sharma P, Sampliner RE, Camargo E. Normalization of esophageal pH with high-dose proton pump inhibitor therapy does not result in regression of Barrett's esophagus. *Am. J. Gastroenterol.* 1997; **92**: 582–5.
82 Rees JR, Lao-Sirieix P, Wong A, Fitzgerald RC. Treatment for Barrett's oesophagus. *Cochrane Database Syst. Rev.* 2010; (1) Cd004060.
83 Ackroyd R, Tam W, Schoeman M, Devitt PG, Watson DI. Prospective randomized controlled trial of argon plasma coagulation ablation vs. endoscopic surveillance of patients with Barrett's esophagus after antireflux surgery. *Gastrointest. Endosc.* 2004; **59**: 1–7.
84 Bright T, Watson DI, Tam W *et al*. Prospective randomized trial of argon plasma coagulation ablation versus endoscopic surveillance of Barrett's esophagus in patients treated with antisecretory medication. *Dig. Dis. Sci.* 2009; **54**: 2606–11.
85 Bright T, Watson DI, Tam W *et al*. Randomized trial of argon plasma coagulation versus endoscopic surveillance for Barrett esophagus after antireflux surgery: late results. *Ann. Surg.* 2007; **246**: 1016–20.
86 Jankowski J. *A Phase III, Randomized, Study of Aspirin and Esomeprazole Chemoprevention in Barrett's Metaplasia (AspECT)*. NCT00357682. Cited 17 February 2015. Available from URL: https://clinicaltrials.gov/show/NCT00357682
87 Overholt BF, Lightdale CJ, Wang KK *et al*. Photodynamic therapy with porfimer sodium for ablation of high-grade dysplasia in Barrett's esophagus: international, partially blinded, randomized phase III trial. *Gastrointest. Endosc.* 2005; **62**: 488–98.
88 Phoa KN, van Vilsteren FG, Weusten BL *et al*. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *JAMA* 2014; **311**: 1209–17.
89 Shaheen NJ, Overholt BF, Sampliner RE *et al*. Durability of radiofrequency ablation in Barrett's esophagus with dysplasia. *Gastroenterology* 2011; **141**: 460–8.
90 Shaheen NJ, Sharma P, Overholt BF *et al*. Radiofrequency ablation in Barrett's esophagus with dysplasia. *N. Engl. J. Med.* 2009; **360**: 2277–88.
91 Sie C, Bright T, Schoeman M *et al*. Argon plasma coagulation ablation versus endoscopic surveillance of Barrett's esophagus: late outcomes from two randomized trials. *Endoscopy* 2013; **45**: 859–65.
92 Cooper GS, Kou TD, Chak A. Receipt of previous diagnoses and endoscopy and outcome from esophageal adenocarcinoma: a population-based study with temporal trends. *Am. J. Gastroenterol.* 2009; **104**: 1356–62.
93 Rubenstein JH, Sonnenberg A, Davis J, McMahon L, Inadomi JM. Effect of a prior endoscopy on outcomes of esophageal adenocarcinoma among United States veterans. *Gastrointest. Endosc.* 2008; **68**: 849–55.
94 Fitzgerald RC, Saeed IT, Khoo D, Farthing MJ, Burnham WR. Rigorous surveillance protocol increases detection of curable cancers associated with Barrett's esophagus. *Dig. Dis. Sci.* 2001; **46**: 1892–8.
95 Reid BJ, Blount PL, Feng Z, Levine DS. Optimizing endoscopic biopsy detection of early cancers in Barrett's high-grade dysplasia. *Am. J. Gastroenterol.* 2000; **95**: 3089–96.
96 Ramus JR, Gatenby PA, Caygill CP, Winslet MC, Watson A. Surveillance of Barrett's columnar-lined oesophagus in the UK: endoscopic intervals and frequency of detection of dysplasia. *Eur. J. Gastroenterol. Hepatol.* 2009; **21**: 636–41.
97 Abrams JA, Kapel RC, Lindberg GM *et al*. Adherence to biopsy guidelines for Barrett's esophagus surveillance in the community setting in the United States. *Clin. Gastroenterol. Hepatol.* 2009; **7**: 736–42.
98 Curvers WL, ten Kate FJ, Krishnadath KK *et al*. Low-grade dysplasia in Barrett's esophagus: overdiagnosed and underestimated. *Am. J. Gastroenterol.* 2010; **105**: 1523–30.
99 Hirst NG, Gordon LG, Whiteman DC, Watson DI, Barendregt JJ. Is endoscopic surveillance for non-dysplastic Barrett's esophagus cost-effective? Review of economic evaluations. *J. Gastroenterol. Hepatol.* 2011; **26**: 247–54.
100 Das A, Wells C, Kim HJ, Fleischer DE, Crowell MD, Sharma VK. An economic analysis of endoscopic ablative therapy for management of nondysplastic Barrett's esophagus. *Endoscopy* 2009; **41**: 400–8.
101 Garside R, Pitt M, Somerville M, Stein K, Price A, Gilbert N. Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling. *Health. Technol. Assess.* 2006; **10**: 1–142.
102 Inadomi JM, Sampliner R, Lagergren J, Lieberman D, Fendrick AM, Vakil N. Screening and surveillance for Barrett esophagus in high-risk groups: a cost-utility analysis. *Ann. Intern. Med.* 2003; **138**: 176–86.
103 Inadomi JM, Somsouk M, Madanick RD, Thomas JP, Shaheen NJ. A cost-utility analysis of ablative therapy for Barrett's esophagus. *Gastroenterology* 2009; **136**: 2101–14.
104 Provenzale D, Schmitt C, Wong JB. Barrett's esophagus: a new look at surveillance based on emerging estimates of cancer risk. *Am. J. Gastroenterol.* 1999; **94**: 2043–53.
105 Sonnenberg A, Fennerty MB. Medical decision analysis of chemoprevention against esophageal adenocarcinoma. *Gastroenterology* 2003; **124**: 1758–66.
106 Sonnenberg A, Soni A, Sampliner RE. Medical decision analysis of endoscopic surveillance of Barrett's oesophagus to prevent oesophageal adenocarcinoma. *Aliment. Pharmacol. Ther.* 2002; **16**: 41–50.
107 Yousef F, Cardwell C, Cantwell MM, Galway K, Johnston BT, Murray L. The incidence of esophageal cancer and high-grade dysplasia in Barrett's esophagus: a systematic review and meta-analysis. *Am. J. Epidemiol.* 2008; **168**: 237–49.
108 Anaparthy R, Gaddam S, Kanakadandi V *et al*. Association between length of Barrett's esophagus and risk of high-grade dysplasia or adenocarcinoma in patients without dysplasia. *Clin. Gastroenterol. Hepatol.* 2013; **11**: 1430–6.
109 Rugge M, Zaninotto G, Parente P *et al*. Barrett's esophagus and adenocarcinoma risk: the experience of the North-Eastern Italian Registry (EBRA). *Ann. Surg.* 2012; **256**: 788–94.

110 Sikkema M, Looman CW, Steyerberg EW *et al*. Predictors for neoplastic progression in patients with Barrett's esophagus: a prospective cohort study. *Am. J. Gastroenterol.* 2011; **106**: 1231–8.
111 Gatenby PA, Caygill CP, Ramus JR, Charlett A, Watson A. Barrett's columnar-lined oesophagus: demographic and lifestyle associations and adenocarcinoma risk. *Dig. Dis. Sci.* 2008; **53**: 1175–85.
112 Verbeek RE, van Oijen MG, ten Kate FJ *et al*. Surveillance and follow-up strategies in patients with high-grade dysplasia in Barrett's esophagus: a Dutch population-based study. *Am. J. Gastroenterol.* 2012; **107**: 534–42.
113 Mandal A, Playford RJ, Wicks AC. Current practice in surveillance strategy for patients with Barrett's oesophagus in the UK. *Aliment. Pharmacol. Ther.* 2003; **17**: 1319–24.
114 Bennett C, Vakil N, Bergman J *et al*. Consensus statements for management of Barrett's dysplasia and early-stage esophageal adenocarcinoma, based on a Delphi process. *Gastroenterology* 2012; **143**: 336–46.
115 Cameron AJ, Carpenter HA. Barrett's esophagus, high-grade dysplasia, and early adenocarcinoma: a pathological study. *Am. J. Gastroenterol.* 1997; **92**: 586–91.
116 Theisen J, Stein HJ, Feith M *et al*. Preferred location for the development of esophageal adenocarcinoma within a segment of intestinal metaplasia. *Surg. Endosc.* 2006; **20**: 235–8.
117 Kariyawasam VC, Bourke MJ, Hourigan LF *et al*. Circumferential location predicts the risk of high-grade dysplasia and early adenocarcinoma in short-segment Barrett's esophagus. *Gastrointest. Endosc.* 2012; **75**: 938–44.
118 Odze RD. Diagnosis and grading of dysplasia in Barrett's oesophagus. *J. Clin. Pathol.* 2006; **59**: 1029–38.
119 Voltaggio L, Montgomery EA, Lam-Himlin D. A clinical and histopathologic focus on Barrett esophagus and Barrett-related dysplasia. *Arch. Pathol. Lab. Med.* 2011; **135**: 1249–60.
120 Lomo LC, Blount PL, Sanchez CA *et al*. Crypt dysplasia with surface maturation: a clinical, pathologic, and molecular study of a Barrett's esophagus cohort. *Am. J. Surg. Pathol.* 2006; **30**: 423–35.
121 Brown IS, Whiteman DC, Lauwers GY. Foveolar type dysplasia in Barrett esophagus. *Mod. Pathol.* 2010; **23**: 834–43.
122 Rugge M, Correa P, Dixon MF *et al*. Gastric dysplasia: the Padova international classification. *Am. J. Surg. Pathol.* 2000; **24**: 167–76.
123 Schlemper RJ, Riddell RH, Kato Y *et al*. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000; **47**: 251–5.
124 Anaparthy R, Sharma P. Progression of Barrett oesophagus: role of endoscopic and histological predictors. *Nat. Rev. Gastroenterol. Hepatol.* 2014; **11**: 525–34.
125 Kerkhof M, van Dekken H, Steyerberg EW *et al*. Grading of dysplasia in Barrett's oesophagus: substantial interobserver variation between general and gastrointestinal pathologists. *Histopathology* 2007; **50**: 920–7.
126 Montgomery E, Bronner MP, Goldblum JR *et al*. Reproducibility of the diagnosis of dysplasia in Barrett esophagus: a reaffirmation. *Hum. Pathol.* 2001; **32**: 368–78.
127 Reid BJ, Haggitt RC, Rubin CE *et al*. Observer variation in the diagnosis of dysplasia in Barrett's esophagus. *Hum. Pathol.* 1988; **19**: 166–78.
128 Duits LC, Phoa KN, Curvers WL *et al*. Barrett's oesophagus patients with low-grade dysplasia can be accurately risk-stratified after histological review by an expert pathology panel. *Gut* 2014; doi:10.1136/gutjnl-2014-307278.
129 Appelman HD. Adenocarcinoma in Barrett mucosa treated by endoscopic mucosal resection. *Arch. Pathol. Lab. Med.* 2009; **133**: 1793–7.
130 Goldblum JR. Controversies in the diagnosis of Barrett esophagus and Barrett-related dysplasia: one pathologist's perspective. *Arch. Pathol. Lab. Med.* 2010; **134**: 1479–84.
131 Downs-Kelly E, Mendelin JE, Bennett AE *et al*. Poor interobserver agreement in the distinction of high-grade dysplasia and adenocarcinoma in pretreatment Barrett's esophagus biopsies. *Am. J. Gastroenterol.* 2008; **103**: 2333–40.
132 Zhu W, Appelman HD, Greenson JK *et al*. A histologically defined subset of high-grade dysplasia in Barrett mucosa is predictive of associated carcinoma. *Am. J. Clin. Pathol.* 2009; **132**: 94–100.
133 Kumarasinghe M, Brown I, Raftopoulos S *et al*. Standardised reporting protocol for endoscopic resection for Barrett oesophagus associated neoplasia: expert consensus recommendations. *Pathol. J. RCPA* 2014; **46**: 473–80.
134 Ayers K, Shi C, Washington K, Yachimski P. Expert pathology review and endoscopic mucosal resection alters the diagnosis of patients referred to undergo therapy for Barrett's esophagus. *Surg. Endosc.* 2013; **27**: 2836–40.
135 Conio M, Repici A, Cestari R *et al*. Endoscopic mucosal resection for high-grade dysplasia and intramucosal carcinoma in Barrett's esophagus: an Italian experience. *World J. Gastroenterol.* 2005; **11**: 6650–5.
136 Mino-Kenudson M, Brugge WR, Puricelli WP *et al*. Management of superficial Barrett's epithelium-related neoplasms by endoscopic mucosal resection: clinicopathologic analysis of 27 cases. *Am. J. Surg. Pathol.* 2005; **29**: 680–6.
137 Wani S, Abrams J, Edmundowicz SA *et al*. Endoscopic mucosal resection results in change of histologic diagnosis in Barrett's esophagus patients with visible and flat neoplasia: a multicenter cohort study. *Dig. Dis. Sci.* 2013; **58**: 1703–9.
138 Larghi A, Lightdale CJ, Memeo L, Bhagat G, Okpara N, Rotterdam H. EUS followed by EMR for staging of high-grade dysplasia and early cancer in Barrett's esophagus. *Gastrointest. Endosc.* 2005; **62**: 16–23.
139 Moss A, Bourke MJ, Hourigan LF *et al*. Endoscopic resection for Barrett's high-grade dysplasia and early esophageal adenocarcinoma: an essential staging procedure with long-term therapeutic benefit. *Am. J. Gastroenterol.* 2010; **105**: 1276–83.
140 Nurkin SJ, Nava HR, Yendamuri S *et al*. Outcomes of endoscopic resection for high-grade dysplasia and esophageal cancer. *Surg. Endosc.* 2014; **28**: 1090–5.
141 Chemaly M, Scalone O, Durivage G *et al*. Miniprobe EUS in the pretherapeutic assessment of early esophageal neoplasia. *Endoscopy* 2008; **40**: 2–6.
142 Young PE, Gentry AB, Acosta RD, Greenwald BD, Riddle M. Endoscopic ultrasound does not accurately stage early adenocarcinoma or high-grade dysplasia of the esophagus. *Clin. Gastroenterol. Hepatol.* 2010; **8**: 1037–41.
143 Pech O, May A, Gunter E, Gossner L, Ell C. The impact of endoscopic ultrasound and computed tomography on the TNM staging of early cancer in Barrett's esophagus. *Am. J. Gastroenterol.* 2006; **101**: 2223–9.
144 Vazquez-Sequeiros E, Wiersema MJ, Clain JE *et al*. Impact of lymph node staging on therapy of esophageal carcinoma. *Gastroenterology* 2003; **125**: 1626–35.
145 von Rahden BH, Stein HJ, Weber A *et al*. Critical reappraisal of current surveillance strategies for Barrett's esophagus: analysis of a large German Barrett's database. *Dis. Esophagus* 2008; **21**: 685–9.
146 Wani S, Puli SR, Shaheen NJ *et al*. Esophageal adenocarcinoma in Barrett's esophagus after endoscopic ablative therapy: a meta-analysis and systematic review. *Am. J. Gastroenterol.* 2009; **104**: 502–13.

147 Paris T. Endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest. Endosc.* 2003; **58** (6 Suppl.): S3–43.
148 Dunbar KB, Spechler SJ. The risk of lymph-node metastases in patients with high-grade dysplasia or intramucosal carcinoma in Barrett's esophagus: a systematic review. *Am. J. Gastroenterol.* 2012; **107**: 850–62.
149 Griffin SM, Burt AD, Jennings NA. Lymph node metastasis in early esophageal adenocarcinoma. *Ann. Surg.* 2011; **254**: 731–6.
150 Leers JM, DeMeester SR, Oezcelik A *et al*. The prevalence of lymph node metastases in patients with T1 esophageal adenocarcinoma a retrospective review of esophagectomy specimens. *Ann. Surg.* 2011; **253**: 271–8.
151 Sepesi B, Watson TJ, Zhou D *et al*. Are endoscopic therapies appropriate for superficial submucosal esophageal adenocarcinoma? An analysis of esophagectomy specimens. *J. Am. Coll. Surg.* 2010; **210**: 418–27.
152 Pohl H, Sonnenberg A, Strobel S, Eckardt A, Rosch T. Endoscopic versus surgical therapy for early cancer in Barrett's esophagus: a decision analysis. *Gastrointest. Endosc.* 2009; **70**: 623–31.
153 Pech O, Bollschweiler E, Manner H, Leers J, Ell C, Holscher AH. Comparison between endoscopic and surgical resection of mucosal esophageal adenocarcinoma in Barrett's esophagus at two high-volume centers. *Ann. Surg.* 2011; **254**: 67–72.
154 Pouw RE, van Vilsteren FG, Peters FP *et al*. Randomized trial on endoscopic resection-cap versus multiband mucosectomy for piecemeal endoscopic resection of early Barrett's neoplasia. *Gastrointest. Endosc.* 2011; **74**: 35–43.
155 Ell C, May A, Pech O *et al*. Curative endoscopic resection of early esophageal adenocarcinomas (Barrett's cancer). *Gastrointest. Endosc.* 2007; **65**: 3–10.
156 Pacifico RJ, Wang KK, Wongkeesong LM, Buttar NS, Lutzke LS. Combined endoscopic mucosal resection and photodynamic therapy versus esophagectomy for management of early adenocarcinoma in Barrett's esophagus. *Clin. Gastroenterol. Hepatol.* 2003; **1**: 252–7.
157 Tian J, Prasad GA, Lutzke LS, Lewis JT, Wang KK. Outcomes of T1b esophageal adenocarcinoma patients. *Gastrointest. Endosc.* 2011; **74**: 1201–6.
158 Bennett C, Green S, Decaestecker J *et al*. Surgery versus radical endotherapies for early cancer and high-grade dysplasia in Barrett's oesophagus. *Cochrane Database Syst. Rev.* 2012; (11) Cd007334.
159 Menon D, Stafinski T, Wu H, Lau D, Wong C. Endoscopic treatments for Barrett's esophagus: a systematic review of safety and effectiveness compared to esophagectomy. *BMC Gastroenterol.* 2010; **10**: 111.
160 Prasad GA, Wang KK, Buttar NS *et al*. Long-term survival following endoscopic and surgical treatment of high-grade dysplasia in Barrett's esophagus. *Gastroenterology* 2007; **132**: 1226–33.
161 Schembre DB, Huang JL, Lin OS, Cantone N, Low DE. Treatment of Barrett's esophagus with early neoplasia: a comparison of endoscopic therapy and esophagectomy. *Gastrointest. Endosc.* 2008; **67**: 595–601.
162 Wu J, Pan YM, Wang TT, Gao DJ, Hu B. Endotherapy versus surgery for early neoplasia in Barrett's esophagus: a meta-analysis. *Gastrointest. Endosc.* 2014; **79**: 233–41.e2.

FULL PAPER

BJC

British Journal of Cancer (2015) 112, 1384–1391 | doi: 10.1038/bjc.2015.100

Keywords: Barrett's oesophagus; intestinal metaplasia; oesophageal adenocarcinoma; progression; MIC-1; GDF15; prognostic biomarker

MIC-1/GDF15 in Barrett's oesophagus and oesophageal adenocarcinoma

O M Fisher[1], A J Levert-Mignon[1], S J Lord[1,2,3], K K M Lee-Ng[1], N K Botelho[1], D Falkenback[1,4], M L Thomas[1], Y V Bobryshev[1,5], D C Whiteman[6], D A Brown[1,7], S N Breit[1] and R V Lord*[,1,8]

[1]St Vincent's Centre for Applied Medical Research and University of New South Wales, Sydney, NSW 2010 Australia; [2]NHMRC Clinical Trials Centre University of Sydney, Sydney, NSW 2050, Australia; [3]Department of Epidemiology and Medical Statistics, School of Medicine, University of Notre Dame, Sydney, NSW 2010 Australia; [4]Department of Surgery, Lund University Hospital (Skåne University Hospital) and Lund University, Lund 221 85, Sweden; [5]Faculty of Medicine, School of Medical Sciences, University of New South Wales, Sydney, Australia; [6]QIMR Berghofer Medical Research Institute, Brisbane, Australia; [7]Peter Duncan Neuroscience Research Unit, St Vincent's Centre for Applied Medical Research, Sydney, NSW 2010 Australia and [8]Department of Surgery, School of Medicine, University of Notre Dame, Sydney, NSW 2010 Australia

Background: Biomarkers are needed to improve current diagnosis and surveillance strategies for patients with Barrett's oesophagus (BO) and oesophageal adenocarcinoma (OAC). Macrophage inhibitory cytokine 1/growth differentiation factor 15 (MIC-1/GDF15) tissue and plasma levels have been shown to predict disease progression in other cancer types and was therefore evaluated in BO/OAC.

Methods: One hundred thirty-eight patients were studied: 45 normal oesophagus (NE), 37 BO, 16 BO with low-grade dysplasia (LGD) and 40 OAC.

Results: Median tissue expression of MIC-1/GDF15 mRNA was $\geqslant$25-fold higher in BO and LGD compared to NE ($P<0.001$); two-fold higher in OAC *vs* BO ($P=0.039$); and 47-fold higher in OAC *vs* NE ($P<0.001$). Relative MIC-1/GDF15 tissue expression >720 discriminated between the presence of either OAC or LGD *vs* NE with 94% sensitivity and 71% specificity (ROC AUC 0.86, 95% CI 0.73–0.96; $P<0.001$). Macrophage inhibitory cytokine 1/growth differentiation factor 15 plasma values were also elevated in patients with OAC *vs* NE ($P<0.001$) or BO ($P=0.015$). High MIC-1/GDF15 plasma levels ($\geqslant 1140\,\mathrm{pg\,ml^{-1}}$) were an independent predictor of poor survival for patients with OAC (HR 3.87, 95% CI 1.01–14.75; $P=0.047$).

Conclusions: Plasma and tissue levels of MIC-1/GDF15 are significantly elevated in patients with BO, LGD and OAC. Plasma MIC-1/GDF15 may have value in diagnosis and monitoring of Barrett's disease.

Barrett's oesophagus (BO) is an acquired condition in which the normal squamous lining of the distal oesophagus is replaced by a specialised intestinal metaplastic (IM) columnar epithelium (Phillips *et al*, 2011). Barrett's oesophagus is found in ~6–12% of all upper gastrointestinal endoscopies (Ford *et al*, 2005) and in an estimated 1.6% of individuals in Western populations. Known risk factors for BO include gastroesophageal reflux disease (GERD), male sex, age over 50 years, Caucasian ethnicity, obesity (especially with a central/visceral fat distribution) and smoking (Cook *et al*, 2012; Fitzgerald *et al*, 2014).

Barrett's oesophagus is a multistage disease in which a minority of patients progress from IM through the stages of low-grade dysplasia (LGD) to high-grade dysplasia (HGD) and to oesophageal adenocarcinoma (OAC; Clemons *et al*, 2013). Recent population-based data indicate that patients with BO without dysplasia have an ~0.5% annual risk of progression from BO to OAC (Hvid-Jensen *et al*, 2011; Desai *et al*, 2012). This risk is higher for patients with HGD, who may have a 6% or higher annual risk of progressing to cancer (Spechler, 2013). Oesophageal adenocarcinoma is a highly fatal cancer, which has

*Correspondence: Professor RV Lord; E-mail: rvlord@stvincents.com.au

Received 17 December 2014; revised 5 February 2015; accepted 16 February 2015; published online 17 March 2015

increased six-fold in incidence over the past three decades (Eheman *et al*, 2012).

Recent data indicate that endoscopic surveillance may correlate with earlier stage diagnosis and improved survival from cancer (Bhat *et al*, 2014). For this reason, guidelines on the management of BO (Spechler *et al*, 2011; Fitzgerald *et al*, 2014) generally recommend endoscopic surveillance with histopathological assessment of dysplasia in four-quadrant random biopsies taken every 2 cm, in addition to targeted biopsies taken from macroscopically visible lesions (so-called Seattle protocol).

There are significant problems with the endoscopic surveillance of patients with BO, including sampling error and variation in the histopathological interpretation of the degree of dysplasia (Reid *et al*, 1988; Lao-Sirieix and Fitzgerald, 2012). In addition, OAC can also develop in the interval between surveillance endoscopies (Spechler, 2007). Further, only 5–7% of patients presenting with OAC have a previous diagnosis of BO (Dulai *et al*, 2002; Bhat *et al*, 2014).

These deficiencies in current management have stimulated a search for biomarkers to improve both the early detection of OAC and the identification of patients with BO who are at high risk of progressing OAC. So far no biomarkers have proven adequate for routine clinical practice (Varghese *et al*, 2012; Fitzgerald *et al*, 2014). A blood biomarker would have several advantages over the current subjective histologic interpretation of endoscopic tissue biopsies, including being less invasive and safer, less expensive and potentially applicable for at-risk population screening.

Macrophage inhibitory cytokine 1 (MIC-1, also known as growth differentiation factor 15, GDF15), is a divergent member of the transforming growth factor-β (TGF-β) superfamily, which is not highly expressed under normal conditions, other than in the placenta (Bootcov *et al*, 1997; Fairlie *et al*, 1999). Its expression is increased by injury, inflammation or malignancy and it is involved in the pathogenesis of a number of disease including cancer and cardiovascular diseases (Breit *et al*, 2011).

Macrophage inhibitory cytokine 1/growth differentiation factor 15 expression is increased in most cancers including those of the prostate, colon, ovary and breast (Welsh *et al*, 2003; de Wit *et al*, 2005), and various cancer cell lines are known to secrete large amounts of MIC-1/GDF15 (Bauskin *et al*, 2006; Unsicker *et al*, 2013). In many cancers, MIC-1/GDF15 serum levels are associated with histopathological cancer grade, stage and extent of disease, and have been reported as predictors of disease progression in prostate, ovary and colorectal cancer (Bauskin *et al*, 2006). Macrophage inhibitory cytokine 1/growth differentiation factor 15 may also have a role as a clinical biomarker in multiple myeloma, oral squamous cancer and bladder cancer (Brown *et al*, 2003, 2006; Costa *et al*, 2010; Wallin *et al*, 2011; Brown *et al*, 2012; Corre *et al*, 2012). Serum levels of MIC-1/GDF15 are reported to increase during the progression of colorectal cancer (Brown *et al*, 2003) and serum MIC-1/GDF15 measurement is a validated prospective biomarker of the presence of colorectal polyps and cancer, indicating that it has potential as a screening tool in these diseases (Brown *et al*, 2012).

The above data suggest that MIC-1/GDF15 may be useful in the management of BO and OAC. We investigated MIC-1/GDF15 plasma and tissue levels in patients with different stages of BO to assess the potential of MIC-1/GDF15 quantification as a biomarker for diagnosis, for prediction of progression in patients with BO, as well as for diagnosis and prognosis in patients with OAC.

MATERIALS AND METHODS

Study design, study population and specimen collection. We performed a retrospective diagnostic case–control analysis to examine the associations between (a) MIC-1/GDF15 tissue mRNA expression and normal oesophagus (NE), BO, LGD and OAC; (b) MIC-1/GDF15 plasma protein levels and NE, BO, LGD and OAC; and (c) to test the association between MIC-1/GDF15 plasma levels and overall survival in patients with OAC. An analysis of the performance of MIC-1/GDF15 tissue and plasma levels to distinguish between NE, BO, LGD and OAC was also conducted.

The NE, BO and LGD tissues and blood samples were collected at endoscopies performed at St Vincent's Hospital, Sydney, Australia. These were obtained from patients prospectively enroled in an Australia-wide research collaboration entitled PROBE-NET: Progression of Barrett's Esophagus to Esophageal Adenocarcinoma Network. The OAC specimens were either obtained from patients at St Vincent's Hospital or from patients who had been enroled in the population-based case–control Australian Cancer Study through the QIMR Berghofer Medical Research Institute (QIMRB; Whiteman *et al*, 2008). All tissues were formalin fixed and paraffin embedded. The pathological diagnoses were established by pathologists at the respective host institutions. Sections from tissue blocks for mRNA extraction were chosen after reviewing the histopathologic reports and haematoxylin and eosin (H&E) stained slides to confirm the presence of the correct diagnoses. Barrett's oesophagus was diagnosed when there was any length of columnar mucosa in the tubular oesophagus, with IM containing goblet cells on histopathological examination.

Patient plasma samples were collected at study recruitment, centrifuged at 1800 **g**, and the resultant plasma stored at $-80\,^{\circ}C$ until further use. For the analysis of plasma MIC-1/GDF15 levels as a prognostic marker for OAC survival, we used pretreatment samples from an independent cohort of 23 patients with OAC from the Australian Cancer Study (Whiteman *et al*, 2008) and 7 patients treated at St Vincent's Hospital, Sydney. All subjects were treated by oesophagectomy with curative intent and received no preoperative chemotherapy or radiotherapy. Only patients with complete clinicopathological and follow-up data ($n = 27$) were included in the survival analysis.

Institutional review board approval for this study was obtained at all collaborating institutions and all patients provided written informed consent.

RNA isolation. All tissues processed were cut from formalin-fixed, paraffin-embedded tissue blocks. Two 7-μm sections were cut for RNA extraction, which was performed using the Qiagen FFPE RNeasy kit, (cat # 74404, Qiagen, Valencia, CA, USA) following the manufacturer's protocol. Satisfactory RNA yield and quality were confirmed using the BioSpec-nano spectrophotometer (Shimadzu Scientific Instruments, Sydney, Australia).

Multiplexed tandem PCR. Multiplexed tandem PCR (MT-PCR) was used to quantitate the mRNA expression levels of MIC-1/GDF15 and a reference gene, NONO ('non-POU domain-containing, octamer-binding (NONO), transcript variant 2'; NM_007363), using the Rotor-Gene 6000 real-time quantitative PCR system (Corbett Life Sciences/Qiagen, Sydney, Australia), as previously described (Stanley and Szewczuk, 2005; Botelho *et al*, 2010). Briefly, MT-PCR is performed in two stages. In the first stage, isolated RNA is converted into cDNA and amplified using gene-specific primers ('outer' primers). In the second step, the product from stage one is used as a template for PCRs run in a 72-well disc-containing single-gene primers ('inner' primers) in each well. Primers were designed using Primer 3 software modified by AusDiagnostics Pty. Ltd. (AusDiagnostics, Alexandria, New South Wales, Australia). All primer pairs spanned an intron–exon boundary and all samples were run in duplicate. Correct product size and integrity was verified on a Bioanalyzer DNA separation chip (Agilent Technologies, Forest Hill, Victoria, Australia). The relative mRNA expression values were calculated as the ratio of the

raw MIC-1/GDF15 mRNA values to the control gene NONO, with the expression of NONO set to a fixed level (1000).

MIC-1/GDF15 enzyme-linked immunosorbent assay. Plasma MIC-1/GDF15 levels were measured using a validated sandwich enzyme immunoassay, as described previously (Moore *et al*, 2000; Fairlie *et al*, 2001; Brown *et al*, 2002). Briefly, mouse mAb 26G6H6 was used for antigen capture and sheep PAb 233-P was used for detection. The human MIC-1/GDF15 (hMIC-1/GDF15) plasma concentration was determined by reference to a standard curve, which was constructed using recombinant hMIC-1/GDF15. All samples were run in duplicate. To ensure reproducibility, the coefficient of variation for all readings was below 10%. Assay performance was further monitored using standard diagnostic laboratory quality control procedures.

Statistical analysis. Continuous variables were compared using *t*-test statistics, Wilcoxon rank-sum test, one-way analysis of variance and the Kruskal–Wallis test where appropriate. Where necessary, log2 transformation of mRNA relative expression and plasma ELISA values was performed to achieve normal distribution for these analyses. Differences between proportions derived from categorical data were analysed using Pearson's chi-squared test or Fisher's exact test.

The area under the curve (AUC) of the receiver operator characteristic (ROC) curve and 95% confidence interval (CI) was calculated to explore the performance of MIC-1/GDF15 tissue and plasma levels to discriminate between patient pathology groups for three distinct potential clinical uses as follows: (i) to explore the potential of MIC-1/GDF15 tissue and/or plasma measurement as a screening test for OAC, we assessed discrimination between OAC *vs* non-OAC patients (NE, BO and LGD); (ii) to assess the potential of MIC-1/GDF15 measurement as a triage test to rule out low-risk patients who do not require further intervention or endoscopy surveillance, we assessed discrimination between NE and BO *vs* LGD and OAC; and (iii) to assess the performance of tissue MIC-1/GDF15 gene expression and plasma levels as a diagnostic test to identify patients at high risk of developing OAC, we assessed discrimination between OAC and its respective 'high'-risk population (BO IM + LGD). For each of these potential biomarker purposes, the optimal cut point for MIC-1/GDF15 was selected by the Youden's index (Youden, 1950) to inform future validation studies and the sensitivity, specificity at this cut point was reported with 95% CI.

As MIC-1/GDF15 plasma values are known to correlate with nutritional status and BMI (Johnen *et al*, 2007; Breit *et al*, 2011), BMI was recorded for all patients providing plasma samples, and we performed a separate analysis for patients who were non-obese and not underweight (BMI 18.5–29.9 $kg\,m^{-2}$) as defined by the WHO (WHO, 2014).

The Kaplan–Meier method was used to compare survival times for patients classified by MIC-1/GDF15 cutoff levels and differences in survival times were compared using the log-rank test. A Cox proportional hazards model was used to identify independent factors associated with mortality. Only patients with complete clinicopathological and survival data were included for survival analysis ($n = 27$). Data are presented as mean (s.d.), median (interquartile range (IQR)) and OR/HR (95% CI), where applicable. All *P*-values < 0.05 were regarded as statistically significant. For ROC curve AUC analysis, the *P*-value indicates the probability that an observed sample AUC is found when the true AUC is 0.5 (no difference between groups). If the *P*-value is < 0.05, it can be concluded that the AUC is statistically significantly different from 0.5 providing statistical evidence that the biomarker has the ability to distinguish between the two groups (Youden, 1950; Obuchowski, 2003).

All statistical analyses were performed using R Statistical Packages (R Core Team, 2013) and graphing/plotting was performed in Prism (GraphPad Prism version 6.0c for Mac OS X, GraphPad Software, San Diego, California, USA).

RESULTS

Patients and tissues. Demographic data for the 138 patients included in this study are shown in Table 1. Briefly, 45 patients (33%) who underwent clinically indicated upper endoscopy showing no BO and who had no history of GERD were included as healthy controls and provided normal squamous oesophageal tissue and baseline blood samples. Thirty-seven patients (27%) had histologically confirmed BO with IM but no dysplasia, and 16 patients (12%) had BO with LGD. Of the 40 patients with OAC, most had AJCC stage I–IIB disease (73%), whereas 8 patients were stage III (20%) and 1 was stage IV (3%). Despite chart review the correct tumour stage could not be assessed in two patients (5%) due to incomplete clinical data.

MIC-1/GDF15 tissue mRNA expression analysis. MIC-1/GDF15 was significantly overexpressed in BO, LGD and OAC when compared to normal squamous oesophagus (Figure 1A). Median relative mRNA expression of MIC-1/GDF15 increased 25-fold from NE to BO (44.2 to 1092.0; $P < 0.001$), with median expression levels higher in BO–LGD compared to BO, although this difference was not statistically significant (1092.0 *vs* 1185; $P = 0.43$). Macrophage inhibitory cytokine 1/growth differentiation factor 15 expression was 47 times higher in OAC compared to NE (2249.0 *vs* 44.2; $P < 0.001$), and was also significantly higher in OAC compared to BO IM (2249 *vs* 1092; $P = 0.039$).

The AUC for the discriminative performance of MIC-1/GDF15 gene expression to distinguish between OAC and non-OAC (NE, BO and LGD) was 0.82 (95% CI 0.70–0.93; $P < 0.001$). For this analysis, the optimal cut point for relative MIC-1/GDF15 mRNA expression was 961, which yielded a sensitivity of 88% (95% CI 68–97%) and specificity of 64% (95% CI 44–81%) for the detection of OAC (Table 2).

The AUC for the discriminative performance of MIC-1/GDF15 to rule out LGD or OAC was 0.86 (95% CI 0.73–0.96; $P < 0.001$). For this analysis, the optimal cut point for relative MIC-1/GDF15 mRNA expression was 720, which yielded a sensitivity of 94% (95% CI 79–99%) and specificity of 71% (95% CI 48–89%; Table 2). This result indicates that if MIC-1/GDF15 mRNA expression values are low, the presence of OAC and/or dysplastic Barrett's disease can be ruled out with a 6% false-negative rate (95% CI 1–21%).

The diagnostic performance of tissue MIC-1/GDF15 gene expression to distinguish between OAC and the at-risk population (BO IM, LGD) was AUC 0.70 (95% CI 0.51–0.88; Figure 1B) with a sensitivity of 63% (95% CI 41–81%) and specificity of 77% (95% CI 50–93%) at the optimal cut point of 1768 (Table 2).

MIC-1/GDF15 plasma analysis. Median MIC-1/GDF15 plasma values were significantly higher in patients with OAC compared to healthy controls (1018 *vs* 606 $pg\,ml^{-1}$; $P < 0.001$), patients with BO (1018 *vs* 783 $pg\,ml^{-1}$; $P = 0.015$) and patients with BO with LGD (735 $pg\,ml^{-1}$, $P = 0.027$; Figure 2A).

The discriminative performance of plasma MIC-1/GDF15 levels to distinguish between NE, BO, LGD and OAC is shown in Figure 2B and Table 2. For example, for use as a screening test, the AUC of plasma MIC-1/GDF15 levels for the detection of OAC *vs* non-malignant oesophageal findings (NE, BO and LGD) was 0.75 (95% CI 0.65–0.85; $P < 0.001$) with MIC-1/GDF15 plasma levels above an optimal cut point of 811 $pg\,ml^{-1}$ yielding a sensitivity of 83% (95% CI 65–94%) and specificity of 62% (95% CI 50–74%) for the detection of OAC. However, the AUC for MIC-1/GDF15 plasma measurements as a triage test to rule out LGD or OAC did

Table 1. Clinical and pathologic characteristics of included patients

	All ($n=138$)		mRNA analysis ($n=53$)		Plasma analysis ($n=99$)	
	Number	%	Number	%	Number	%
Gender						
Males	102	74	42	79	73	74
Females	36	26	11	21	26	26
Median age, years (range)	60	51–68	63	55–74	61	51–68
Diagnosis						
Normal oesophagus	45	33	12	23	33	33
Barrett's oesophagus IM	37	27	10	19	27	27
Barrett's oesophagus with dysplasia	16	12	7	13	9	9
Oesophageal adenocarcinoma	40	29	24	45	30	30
TNM (7th edition)						
T1–2	31	78	22	92	23	77
T3–4	7	18	—	—	7	23
N1–3	11	8	3	13	10	33
M+	1	3	—	—	1	3
Unknown T, N or M	2	5	2	8	0	—
Tumour stage (AJCC 7th edition)						
IA–B	14	35	9	38	11	37
IIA	13	33	10	42	9	30
IIB	2	5	2	8	1	3
IIIA–C	8	20	1	4	8	27
IV	1	3	—	—	1	3
Unknown stage	2	5	2	8	0	—
Median survival, months (range)	38	18–51	34	18–48	43	23–56

Abbreviations: AJCC = American Joint Committee on Cancer; IM = intestinal metaplasia; TNM = tumour, node and metastasis.

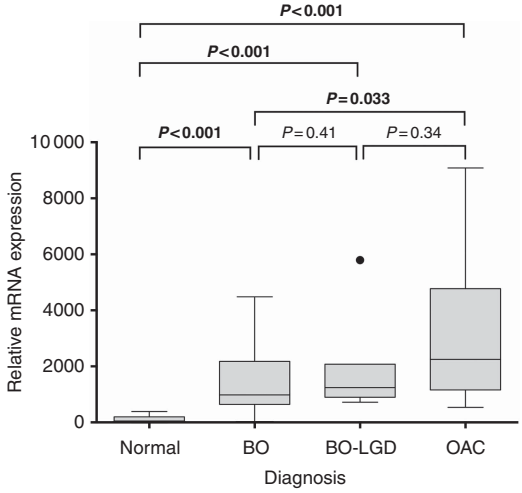


Figure 1. Boxplot of tissue MIC-1/GDF15 mRNA expression analysis by pathology diagnosis. Raw relative mRNA expression values are presented. Differences in relative gene expression values were calculated using Student's *t*-test following log2 transformation. Bold values indicates that these are statistically significant. Dot indicates outlier as is convention for presenting boxplot data.

not indicate potential for discriminating between patient groups (data not shown).

In patients who were non-obese and not underweight (BMI 18.5–29.9 kg m^{-2}) as defined by the WHO (WHO, 2014), median MIC-1/GDF15 plasma levels were slightly lower for all patient groups except for those patients with OAC, where median plasma levels were in a similar range (1040.5 pg ml^{-1}; $P=0.85$). Figure 2C summarises the findings of the plasma analysis when corrected for BMI.

The discriminative performance of plasma MIC-1/GDF15 levels corrected for BMI to distinguish between NE, BO, LGD and OAC is shown in Figure 2D and Table 2. Briefly, non-obese patients with a MIC-1/GDF15 plasma value above the optimal cut point of 811 pg ml^{-1} were approximately four times more likely to have a malignant oesophageal finding than those with lower MIC-1/GDF15 plasma levels (AUC 0.85 (95% CI 0.76–0.95), sensitivity 84% (95% CI 64–95%) and specificity 80% (95% CI 65–90%)).

MIC-1/GDF15 plasma levels and overall survival in OAC. Two patients were excluded from the survival analysis due to incomplete clinicopathological data. Patients were grouped into AJCC stages 1 and 2 in one group and stages 3 and 4 in another group.

There were no significant associations between MIC-1/GDF15 plasma levels and any of the following known prognostic factors: histopathological grade, T-stage, positive nodal status, higher tumour stage (data not shown).

There was a non-significant trend towards worse overall survival in patients with OAC and elevated levels of MIC-1/GDF15 ($\geqslant$1140 pg ml^{-1} (optimal cut point determined by ROC curve analysis); 47.5 *vs* 33.0 months, $P=0.063$; Figure 3). However, in a multivariable Cox proportional hazards model including the independent variables age, BMI, overall tumour stage and plasma MIC-1/GDF15 values $\geqslant$1140 pg ml^{-1}, both elevated plasma MIC-1/GDF15 levels (HR 3.87, 95% CI 1.01–14.75; $P=0.048$) and worse tumour stages (HR 13.85, 95% CI 2.31–83.23, $P=0.004$) were significant independent prognostic markers for mortality (Table 3).

DISCUSSION

In this novel study, we found evidence for an association between MIC-1/GDF15 tissue and plasma levels in patients with BO and OAC. This suggests that MIC-1/GDF15 could be evaluated further for its potential as a biomarker in this disease, with our results indicating a greater biomarker potential for measuring MIC-1/GDF15 in blood rather than tissues. A non-tissue, non-endoscopic biomarker or biomarker panel detectable in blood would offer a

Table 2. Discriminative performance of plasma MIC-1/GDF15 in predicting the presence of oesophageal pathologies

	Youden index[a]	Cutoff point[a]	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	*P*-value[b]
Tissue mRNA ($n=54$)						
Non-malignant *vs* OAC	0.52	961	0.82 (0.70–0.93)	0.88 (0.68–0.97)	0.64 (0.44–0.81)	0.001
NE + BO *vs* LGD + OAC	0.65	720	0.86 (0.73–0.96)	0.94 (0.79–0.99)	0.71 (0.48–0.89)	<0.001
BO + LGD *vs* OAC	0.39	1768	0.70 (0.51–0.88)	0.63 (0.41–0.81)	0.77 (0.50–0.93)	0.03
Plasma ELISA for all patients ($n=99$)						
Non-malignant *vs* OAC	0.46	811	0.75 (0.65–0.85)	0.83 (0.65–0.94)	0.62 (0.50–0.74)	<0.001
NE *vs* BO + LGD + OAC	0.35	705	0.70 (0.58–0.82)	0.74 (0.62–0.84)	0.61 (0.42–0.77)	0.001
BO + LGD *vs* OAC	0.39	811	0.72 (0.56–0.84)	0.83 (0.65–0.94)	0.56 (0.38–0.72)	0.003
Plasma ELISA for patients with a BMI of 18.5–29.9 kg m^{-2} ($n=69$)						
Non-malignant *vs* OAC	0.63	811	0.85 (0.76–0.95)	0.84 (0.64–0.95)	0.80 (0.65–0.90)	<0.001
NE *vs* BO + LGD + OAC	0.44	705	0.77 (0.65–0.90)	0.69 (0.55–0.82)	0.75 (0.51–0.91)	<0.001
BO + LGD *vs* OAC	0.59	836	0.83 (0.71–0.95)	0.76 (0.55–0.91)	0.83 (0.63–0.95)	<0.001

Abbreviations: AUC = area under the curve; BO = Barrett's oesophagus; BMI = body mass index; CI = confidence interval; OAC = oesophageal adenocarcinoma; ELISA = enzyme-linked immunosorbent assay; LGD = BO with low-grade dysplasia; MIC-1/GDF15 = macrophage inhibitory cytokine 1/growth differentiation factor 15; NE = normal oesophagus/healthy controls; ROC = receiver operator characteristic.
[a]Grouped as more *vs* less than the cutoff point generated by the ROC and Youden index; for relative gene expression no units apply for plasma measurements (pg ml^{-1}).
[b]*P*-value <0.05 indicates that observed AUC is significantly different from 0.5 providing statistical evidence that the biomarker has the ability to distinguish between the two groups.

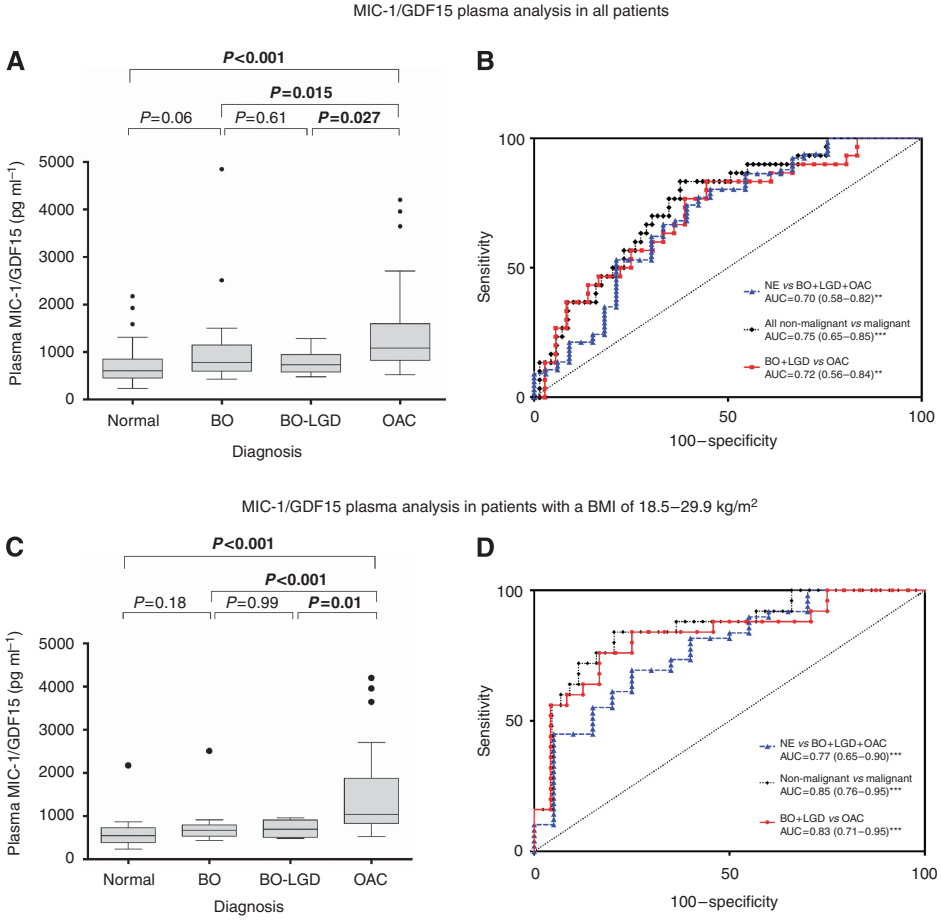


Figure 2. Boxplots of MIC-1/GDF15 plasma levels by pathology diagnosis and corresponding receiver operator characteristic (ROC) curves for performance analysis of tissue MIC-1/GDF15 plasma levels to discriminate patient groups. (**A** and **B**) Show the analysis of the whole study cohort, whereas (**C** and **D**) depict the findings in the non-obese, non-underweight (as categorised by the WHO) patient population. Differences in plasma values were calculated using Student's *t*-test following log2 transformation. Data are presented untransformed as pg ml^{-1}. Bold values indicates that these are statistically significant. Dot indicates outlier as is convention for presenting boxplot data. **$P<0.01$; ***$P<0.001$.

major advantage if it could replace the need for pathology classification of dysplasia. Although the presence of dysplasia in BO detected by histopathological examination of tissue specimens is the most informative current predictor of risk of progression to OAC, it is not a reliable tests as shown by the disturbing lack of agreement between pathologists and their interpretation of presence or severity of dysplasia (Reid *et al*, 1988; Spechler, 2007). Histopathological examination of the primary tumour and

resected lymph nodes along with radiologic tests is also the current basis of staging patients with OAC and determining prognosis, but similarly has limited accuracy. For example, ~50% of patients assessed as having no lymph node involvement at diagnosis, and thus a good prognosis, will still die of their disease by 10 years (Rice *et al*, 2003). Identifying predictive and prognostic biomarkers for BO and OAC is thus crucial to improve clinical care.

Our evaluation of MIC-1/GDF15 plasma quantification as a diagnostic test for OAC found that patients with high MIC-1/GDF15 plasma levels who were neither underweight nor obese were more than four times more likely to have OAC. This likelihood decreased by two-fold if obese or underweight patients were included in the analysis. This reduction may be explained by the confounding effect of the known relationship of MIC-1/GDF15 blood levels and BMI (Johnen *et al*, 2007; Breit *et al*, 2011). This finding addresses a potential limitation of the clinical applicability of plasma MIC-1/GDF15 analysis for the diagnosis and monitoring of patients with Barrett's disease, as patients with BO usually have higher BMI and increased waist circumference (Kubo *et al*, 2013). Conversely, however, as elevated MIC-1/GDF15 serum levels are known to be associated with tumour-induced weight loss (Wakchoure *et al*, 2009; Tsai *et al*, 2012), patients with noted weight loss (despite the lack of dysphagia/reduced food intake) and elevated plasma MIC-1 levels may identify those at highest risk of bearing an oesophageal malignancy.

Others have shown the potential clinical use for MIC-1/GDF15 in blood for the diagnosis and/or monitoring of pancreatic, prostate, colon and thyroid cancers, especially when MIC-1/GDF15 is combined with other markers, as we anticipate would also be the case with OAC (Koopmann *et al*, 2004; Brown *et al*, 2006). In pancreatic cancer, for example, combining MIC-1/GDF15 serum levels with CA19-9 significantly improved the diagnosis of pancreatic cancer leading to a sensitivity of 70% and specificity of 85% (Koopmann *et al*, 2004). In prostate cancer, the combination of serum MIC-1/GDF15 levels with prostate-specific antigen significantly improved the overall diagnostic specificity and shows great potential for monitoring of disease progression (Brown *et al*, 2006).

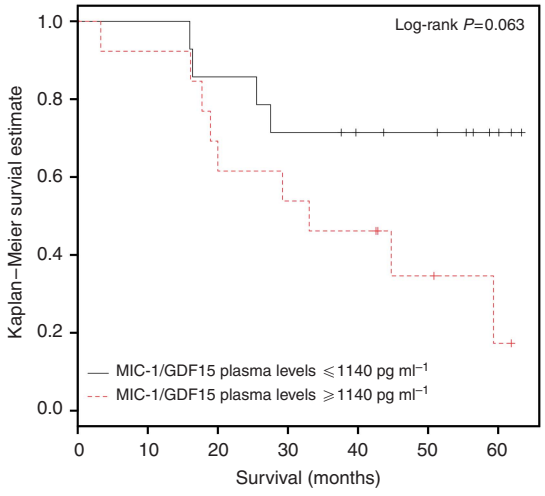


Figure 3. Survival of OAC patients stratified by MIC-1/GDF15 plasma levels $\geqslant 1140\,pg\,ml^{-1}$ (as determined by ROC curve analysis).

In this study, we also found that elevated MIC-1/GDF15 plasma levels were an independent prognostic marker for reduced overall survival in patients, although the reliability of this finding is diminished by the small number of patients in this section ($n = 27$). This finding was significant when MIC-1/GDF15 was included in a multivariable Cox proportional hazards model, which was adjusted for age, tumour stage and BMI. The patients in this study were treated by oesophagectomy alone in all but one case. Because no patient received neoadjuvant chemotherapy or radiotherapy, our findings are unlikely to be explained by variations in treatment or the effect of treatment on tumour biology.

Studies of patients with other malignancies have suggested that MIC-1/GDF15 is a marker of adverse prognosis as its blood levels increased with increasing histopathological grade, invasiveness and metastasis (Mimeault and Batra, 2010; Breit *et al*, 2011). However, in this study, we did not find a significant difference in plasma MIC-1/GDF15 levels between patients with lower *vs* higher stages of OAC. We believe this is most likely due to the limited number of patients with more advanced stages of the disease. Further studies are required to help resolve whether blood levels of MIC-1/GDF15 can help in OAC staging. A blood marker that can identify patients with locally limited but perhaps occult disseminated disease would potentially allow early allocation to more aggressive treatment strategies.

Our finding that elevated plasma MIC-1/GDF15 provides independent prognostic information contrasts with the only other report on MIC-1/GDF15 blood levels that included patients with OAC (Skipworth *et al*, 2010). In this study, patients with elevated MIC-1/GDF15 levels showed reduced survival compared to those with lower MIC-1/GDF15 levels, but MIC-1/GDF15 was not an independent prognostic indicator. However, this study also included patients with oesophageal squamous, undifferentiated and neuroendocrine carcinoma, and 52% of the included cancers were not of oesophageal origin (oesophageal junction with undefined localisation or gastric cancer) leading to a more heterogeneous study population. In contrast to Skipworth *et al*, 2010 however, we did not have data on systemic inflammatory markers and were therefore not able to adjust for this in our Cox regression model.

In oesophageal tissue biopsies, we found that MIC-1/GDF15 mRNA levels were associated with the presence of BO as well as the stage of the disease (IM, LGD and OAC). These findings indicate it may have potential to aid in conventional histopathology methods. This could be explored in future studies by comparing MIC-1/GDF15 results with conventional methods to determine its clinical value for patient management. In particular, given the high clinical need to identify patients at risk of progression to OAC, our results suggest that it may be useful to explore tissue MIC-1/GDF15 levels

Table 3. Univariable and multivariable Cox proportional hazards model for survival ($n = 27$)

	Univariable analysis			Multivariable analysis		
Variable	**HR**	**95% CI**	***P*-value**	**HR**	**95% CI**	***P*-value**
Age	0.99	0.95–1.04	0.82	1.04	0.98–1.11	0.23
Body mass index	0.96	0.84–1.10	0.59	0.99	0.85–1.15	0.88
High tumour stage (III/IV)	5.85	1.92–17.85	0.0019	13.86	2.31–83.23	0.004
MIC-1/GDF15 plasma levels $\geqslant 1140\,pg\,ml^{-1}$	2.91	0.89–9.48	0.076	3.87	1.01–14.75	0.048

Abbreviations: CI = confidence interval; HR = hazard ratio; MIC-1/GDF15 = macrophage inhibitory cytokine 1/growth differentiation factor 15.

in prospective studies of BO patients, in contrast to the cross-sectional design of the present study. It would also be useful to investigate through *in vivo* studies (which we have not performed), the correlation between MIC-1/GDF15 tissue and plasma levels in this disease, as in other malignancies the intracellular processing of MIC-1/GDF15 ultimately determines the relative amount of MIC-1/GDF that remains in the tumour microenvironment *vs* the amount diffusing into the systemic circulation. This additive predictive value has been demonstrated in prostate cancer stromal staining, where detection of proMIC-1/GDF15 was an important independent predictor of disease relapse (Bauskin *et al*, 2005, 2006).

There are two possible explanations for our findings in tissues. First, the increased MIC-1/GDF15 levels in BO IM, BO–LGD and OAC may be linked to the inflammatory response of the oesophagus to severe GERD, as MIC-1/GDF15 is increased in response to multiple cellular stressors and following acute injury and/or inflammatory tissue changes (Fairlie *et al*, 1999; Welsh *et al*, 2003; Unsicker *et al*, 2013). Alternatively (or in addition) MIC-1/GDF15 levels may also signal neoplastic progression of BO to OAC, which is consistent with data in colon cancer, in which MIC-1/GDF15 levels increase with disease progression from normal to adenoma, carcinoma and metastatic disease (Brown *et al*, 2003).

There are clear limitations to our findings. First, due to the relatively small sample size of the present study, the results presented must be interpreted with caution and require confirmation in future, larger well-described Barrett's/OAC patient cohorts. Second, tissue specimen sampling error may occur, a general problem in all studies using endoscopic biopsy specimens from patients with Barrett's disease (Lao-Sirieix and Fitzgerald, 2012). By reviewing the histology of all specimens used prior to RNA extraction, we attempted to minimise non-uniformity in the tissue specimens used. Third, our estimates of sensitivity and specificity are drawn from a selective sample of patients presenting for endoscopy and therefore need to be interpreted with caution due to the potential for spectrum bias, which is associated with overestimates of test accuracy (Pepe *et al*, 2008). Lastly, MIC-1/GDF15 plasma levels are known to be influenced by the use of non-steroidal anti-inflammatory drugs (NSAIDs; Brown *et al*, 2012), which in turn have been shown to influence oesophageal adenocarcinogenesis and the prognosis of patients affected by this cancer (Buskens *et al*, 2002). We did not have complete data on patients' NSAID use and thus were not able to correct for this possible confounding factor.

In conclusion, however, this study provides evidence to suggest that blood MIC-1/GDF15 measurements may have a clinical role for diagnosis and monitoring of Barrett's disease spectrum. These findings need to be confirmed in larger prospective studies, in particular studies including serial MIC-1/GDF15 blood measurements in patients with different stages of BO to determine how dynamic changes in MIC-1/GDF15 levels may be used to inform patient management. Further, for patients already affected by OAC, blood levels of MIC-1/GDF15 may help provide additional prognostic information.

ACKNOWLEDGEMENTS

This work was supported by the National Health and Medical Research Council (NHMRC 1040947), Cancer Council New South Wales (SRP 08-04 and RG 13-03) and St Vincent's Clinic Foundation, Sydney. OMF was supported by the Swiss Cancer League (BIL KLS-3133-02-2013). DF was supported by the Swedish Society of Medicine, the Maggie Stephens Foundation and the Sparre Foundation. DAB is a recipient of a NHMRC Career Development Fellowship.

CONFLICT OF INTEREST

DAB and SNB are named inventors on patents owned by St Vincent's Hospital that pertain to the clinical use of MIC-1/GDF15 diagnostic assay and modulatory therapy. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

Bauskin AR, Brown DA, Junankar S, Rasiah KK, Eggleton S, Hunter M, Liu T, Smith D, Kuffner T, Pankhurst GJ, Johnen H, Russell PJ, Barret W, Stricker PD, Grygiel JJ, Kench JG, Henshall SM, Sutherland RL, Breit SN (2005) The propeptide mediates formation of stromal stores of PROMIC-1: role in determining prostate cancer outcome. *Cancer Res* **65**: 2330–2336.

Bauskin AR, Brown DA, Kuffner T, Johnen H, Luo XW, Hunter M, Breit SN (2006) Role of macrophage inhibitory cytokine-1 in tumorigenesis and diagnosis of cancer. *Cancer Res* **66**: 4983–4986.

Bhat SK, McManus DT, Coleman HG, Johnston BT, Cardwell CR, McMenamin U, Bannon F, Hicks B, Kennedy G, Gavin AT, Murray LJ (2014) Oesophageal adenocarcinoma and prior diagnosis of Barrett's oesophagus: a population-based study. *Gut* **64**: 20–25.

Bootcov MR, Bauskin AR, Valenzuela SM, Moore AG, Bansal M, He XY, Zhang HP, Donnellan M, Mahler S, Pryor K, Walsh BJ, Nicholson RC, Fairlie WD, Por SB, Robbins JM, Breit SN (1997) MIC-1, a novel macrophage inhibitory cytokine, is a divergent member of the TGF-beta superfamily. *Proc Natl Acad Sci USA* **94**: 11514–11519.

Botelho NK, Schneiders FI, Lord SJ, Freeman AK, Tyagi S, Nancarrow DJ, Hayward NK, Whiteman DC, Lord RV (2010) Gene expression alterations in formalin-fixed, paraffin-embedded Barrett esophagus and esophageal adenocarcinoma tissues. *Cancer Biol Ther* **10**: 172–179.

Breit SN, Johnen H, Cook AD, Tsai VW, Mohammad MG, Kuffner T, Zhang HP, Marquis CP, Jiang L, Lockwood G, Lee-Ng M, Husaini Y, Wu L, Hamilton JA, Brown DA (2011) The TGF-beta superfamily cytokine, MIC-1/GDF15: a pleotrophic cytokine with roles in inflammation, cancer and metabolism. *Growth Factors* **29**: 187–195.

Brown DA, Bauskin AR, Fairlie WD, Smith MD, Liu T, Xu N, Breit SN (2002) Antibody-based approach to high-volume genotyping for MIC-1 polymorphism. *Biotechniques* **33**: 118–120, 122, 124 passim.

Brown DA, Hance KW, Rogers CJ, Sansbury LB, Albert PS, Murphy G, Laiyemo AO, Wang Z, Cross AJ, Schatzkin A, Danta M, Srasuebkul P, Amin J, Law M, Breit SN, Lanza E (2012) Serum macrophage inhibitory cytokine-1 (MIC-1/GDF15): a potential screening tool for the prevention of colon cancer? *Cancer Epidemiol Biomarkers Prev* **21**: 337–346.

Brown DA, Stephan C, Ward RL, Law M, Hunter M, Bauskin AR, Amin J, Jung K, Diamandis EP, Hampton GM, Russell PJ, Giles GG, Breit SN (2006) Measurement of serum levels of macrophage inhibitory cytokine 1 combined with prostate-specific antigen improves prostate cancer diagnosis. *Clin Cancer Res* **12**: 89–96.

Brown DA, Ward RL, Buckhaults P, Liu T, Romans KE, Hawkins NJ, Bauskin AR, Kinzler KW, Vogelstein B, Breit SN (2003) MIC-1 serum level and genotype: associations with progress and prognosis of colorectal carcinoma. *Clin Cancer Res* **9**: 2642–2650.

Buskens CJ, Van Rees BP, Sivula A, Reitsma JB, Haglund C, Bosma PJ, Offerhaus GJ, Van Lanschot JJ, Ristimäki A (2002) Prognostic significance of elevated cyclooxygenase 2 expression in patients with adenocarcinoma of the esophagus. *Gastroenterology* **122**: 1800–1807.

Clemons NJ, Phillips WA, Lord RV (2013) Signaling pathways in the molecular pathogenesis of adenocarcinomas of the esophagus and gastresophageal junction. *Cancer Biol Ther* **14**: 782–795.

Cook MB, Shaheen NJ, Anderson LA, Giffen C, Chow WH, Vaughan TL, Whiteman DC, Corley DA (2012) Cigarette smoking increases risk of Barrett's esophagus: an analysis of the Barrett's and Esophageal Adenocarcinoma Consortium. *Gastroenterology* **142**: 744–753.

Corre J, Labat E, Espagnolle N, Hébraud B, Avet-Loiseau H, Roussel M, Huynh A, Gadelorge M, Cordelier P, Klein B, Moreau P, Facon T, Fournié JJ, Attal M, Bourin P (2012) Bioactivity and prognostic significance of

growth differentiation factor GDF15 secreted by bone marrow mesenchymal stem cells in multiple myeloma. *Cancer Res* **72**: 1395–1406.

Costa VL, Henrique R, Danielsen SA, Duarte-Pereira S, Eknaes M, Skotheim RI, Rodrigues A, Magalhaes JS, Oliveira J, Lothe RA, Teixeira MR, Jeronimo C, Lind GE (2010) Three epigenetic biomarkers, GDF15, TMEFF2, and VIM, accurately predict bladder cancer from DNA-based analyses of urine samples. *Clin Cancer Res* **16**: 5842–5851.

de Wit NJ, Rijntjes J, Diepstra JH, van Kuppevelt TH, Weidle UH, Ruiter DJ, van Muijen GN (2005) Analysis of differential gene expression in human melanocytic tumour lesions by custom made oligonucleotide arrays. *Br J Cancer* **92**: 2249–2261.

Desai TK, Krishnan K, Samala N, Singh J, Cluley J, Perla S, Howden CW (2012) The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut* **61**: 970–976.

Dulai GS, Guha S, Kahn KL, Gornbein J, Weinstein WM (2002) Preoperative prevalence of Barrett's esophagus in esophageal adenocarcinoma: a systematic review. *Gastroenterology* **122**: 26–33.

Eheman C, Henley SJ, Ballard-Barbash R, Jacobs EJ, Schymura MJ, Noone AM, Pan L, Anderson RN, Fulton JE, Kohler BA, Jemal A, Ward E, Plescia M, Ries LA, Edwards BK (2012) Annual Report to the Nation on the status of cancer, 1975-2008, featuring cancers associated with excess weight and lack of sufficient physical activity. *Cancer* **118**: 2338–2366.

Fairlie WD, Moore AG, Bauskin AR, Russell PK, Zhang HP, Breit SN (1999) MIC-1 is a novel TGF-beta superfamily cytokine associated with macrophage activation. *J Leukoc Biol* **65**: 2–5.

Fairlie WD, Russell PK, Wu WM, Moore AG, Zhang HP, Brown PK, Bauskin AR, Breit SN (2001) Epitope mapping of the transforming growth factor-beta superfamily protein, macrophage inhibitory cytokine-1 (MIC-1): identification of at least five distinct epitope specificities. *Biochemistry* **40**: 65–73.

Fitzgerald RC, di Pietro M, Ragunath K, Ang Y, Kang JY, Watson P, Trudgill N, Patel P, Kaye PV, Sanders S, O'Donovan M, Bird-Lieberman E, Bhandari P, Jankowski JA, Attwood S, Parsons SL, Loft D, Lagergren J, Moayyedi P, Lyratzopoulos G, de Caestecker J. British Society of Gastroenterology (2014) British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* **63**: 7–42.

Ford AC, Forman D, Reynolds PD, Cooper BT, Moayyedi P (2005) Ethnicity, gender, and socioeconomic status as risk factors for esophagitis and Barrett's esophagus. *Am J Epidemiol* **162**: 454–460.

Hvid-Jensen F, Pedersen L, Drewes AM, Sorensen HT, Funch-Jensen P (2011) Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med* **365**: 1375–1383.

Johnen H, Lin S, Kuffner T, Brown DA, Tsai VW, Bauskin AR, Wu L, Pankhurst G, Jiang L, Junankar S, Hunter M, Fairlie WD, Lee NJ, Enriquez RF, Baldock PA, Corey E, Apple FS, Murakami MM, Lin EJ, Wang C, During MJ, Sainsbury A, Herzog H, Breit SN (2007) Tumor-induced anorexia and weight loss are mediated by the TGF-beta superfamily cytokine MIC-1. *Nat Med* **13**: 1333–1340.

Koopmann J, Buckhaults P, Brown DA, Zahurak ML, Sato N, Fukushima N, Sokoll LJ, Chan DW, Yeo CJ, Hruban RH, Breit SN, Kinzler KW, Vogelstein B, Goggins M (2004) Serum macrophage inhibitory cytokine 1 as a marker of pancreatic and other periampullary cancers. *Clin Cancer Res* **10**: 2386–2392.

Kubo A, Cook MB, Shaheen NJ, Vaughan TL, Whiteman DC, Murray L, Corley DA (2013) Sex-specific associations between body mass index, waist circumference and the risk of Barrett's oesophagus: a pooled analysis from the international BEACON consortium. *Gut* **62**: 1684–1691.

Lao-Sirieix P, Fitzgerald RC (2012) Screening for oesophageal cancer. *Nat Rev Clin Oncol* **9**: 278–287.

Mimeault M, Batra SK (2010) Divergent molecular mechanisms underlying the pleiotropic functions of macrophage inhibitory cytokine-1 in cancer. *J Cell Physiol* **224**: 626–635.

Moore AG, Brown DA, Fairlie WD, Bauskin AR, Brown PK, Munier ML, Russell PK, Salamonsen LA, Wallace EM, Breit SN (2000) The transforming growth factor-ss superfamily cytokine macrophage inhibitory cytokine-1 is present in high concentrations in the serum of pregnant women. *J Clin Endocrinol Metab* **85**: 4781–4788.

Obuchowski NA (2003) Receiver operating characteristic curves and their use in radiology. *Radiology* **229**: 3–8.

Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD (2008) Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* **100**: 1432–1438.

Phillips WA, Lord RV, Nancarrow DJ, Watson DI, Whiteman DC (2011) Barrett's esophagus. *J Gastroenterol Hepatol* **26**: 639–648.

R Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.

Reid BJ, Haggitt RC, Rubin CE, Roth G, Surawicz CM, Van Belle G, Lewin K, Weinstein WM, Antonioli DA, Goldman H (1988) Observer variation in the diagnosis of dysplasia in Barrett's esophagus. *Hum Pathol* **19**: 166–178.

Rice TW, Blackstone EH, Rybicki LA, Adelstein DJ, Murthy SC, DeCamp MM, Goldblum JR (2003) Refining esophageal cancer staging. *J Thorac Cardiovasc Surg* **125**: 1103–1113.

Skipworth RJ, Deans DA, Tan BH, Sangster K, Paterson-Brown S, Brown DA, Hunter M, Breit SN, Ross JA, Fearon KC (2010) Plasma MIC-1 correlates with systemic inflammation but is not an independent determinant of nutritional status or survival in oesophago-gastric cancer. *Br J Cancer* **102**: 665–672.

Spechler SJ (2007) Screening and surveillance for Barrett's esophagus-an unresolved dilemma. *Nat Clin Pract Gastroenterol Hepatol* **4**: 470–471.

Spechler SJ (2013) Barrett's esophagus: the American perspective. *Dig Dis* **31**: 10–16.

Spechler SJ, Sharma P, Souza RF, Inadomi JM, Shaheen NJ (2011) American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* **140**: 1084–1091.

Stanley KK, Szewczuk E (2005) Multiplexed tandem PCR: gene profiling from small amounts of RNA using SYBR Green detection. *Nucleic Acids Res* **33**: e180.

Tsai VW, Husaini Y, Manandhar R, Lee-Ng KK, Zhang HP, Harriott K, Jiang L, Lin S, Sainsbury A, Brown DA, Breit SN (2012) Anorexia/cachexia of chronic diseases: a role for the TGF-beta family cytokine MIC-1/GDF15. *J Cachexia Sarcopenia Muscle* **3**: 239–243.

Unsicker K, Spittau B, Krieglstein K (2013) The multiple facets of the TGF-beta family cytokine growth/differentiation factor-15/macrophage inhibitory cytokine-1. *Cytokine Growth Factor Rev* **24**: 373–384.

Varghese S, Lao-Sirieix P, Fitzgerald RC (2012) Identification and clinical implementation of biomarkers for Barrett's esophagus. *Gastroenterology* **142**: 435–441e2.

WHO (2014) *Global Database on Body Mass Index*, http://apps.who.int/bmi/index.jsp?introPage=intro_3.html Online. http://apps.who.int/bmi/index.jsp?introPage=intro_3.html. Accessed 5 February 2014.

Wakchoure S, Swain TM, Hentunen TA, Bauskin AR, Brown DA, Breit SN, Vuopala KS, Harris KW, Selander KS (2009) Expression of macrophage inhibitory cytokine-1 in prostate cancer bone metastases induces osteoclast activation and weight loss. *Prostate* **69**: 652–661.

Wallin U, Glimelius B, Jirstrom K, Darmanis S, Nong RY, Ponten F, Johansson C, Pahlman L, Birgisson H (2011) Growth differentiation factor 15: a prognostic marker for recurrence in colorectal cancer. *Br J Cancer* **104**: 1619–1627.

Welsh JB, Sapinoso LM, Kern SG, Brown DA, Liu T, Bauskin AR, Ward RL, Hawkins NJ, Quinn DI, Russell PJ, Sutherland RL, Breit SN, Moskaluk CA, Frierson Jr HF, Hampton GM (2003) Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc Natl Acad Sci USA* **100**: 3410–3415.

Whiteman DC, Sadeghi S, Pandeya N, Smithers BM, Gotley DC, Bain CJ, Webb PM, Green AC. Australian Cancer Study (2008) Combined effects of obesity, acid reflux and smoking on the risk of adenocarcinomas of the oesophagus. *Gut* **57**: 173–180.

Youden WJ (1950) Index for rating diagnostic tests. *Cancer* **3**: 32–35.