Title Page

DNA Methylation in Rectal Cancer:
Validating Findings of an Epigenome-Wide Association Study

Submitted by Conor Stephen Jones, to the University of Exeter as a thesis for the degree of Masters by Research in Medical Studies

Date of submission: May 2018

I certify that all material in this thesis/dissertation which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Abstract

Background

Preliminary studies conducted by our group utilised the Illumina Infinium Human Methylation 450k Beadchip array to perform an epigenome-wide association study (EWAS) of 15 matched rectal tumour (RT) and adjacent mucosa (AM) samples. 176 differentially methylated probes (DMPs) were identified (P<0.00001). RT was also characterised by significantly reduced global methylation in comparison to AM.

Aims

This study aimed to validate specific and global DNA methylation differences identified by our preliminary work. We then sought to replicate the findings in additional samples. Finally, we attempted to identify correlations between DNA methylation differences and clinicopathological tumour features.

Materials and Methods

Polymerase chain reaction (PCR) and bisulphite pyrosequencing assays were designed and optimised to quantify DNA methylation at nine DMPs nominated by our EWAS. Pearson's test was used to calculate the correlation between 450k and pyrosequencing methylation values. Replication was performed in an additional cohort of 68 matched colorectal tumour and AM pairs. Global DNA methylation of the discovery cohort was quantified using the luminometric methylation assay (LUMA). Potential relationships between tumour features and differential methylation were investigated using univariate (t-tests or ANOVA) and multivariate analyses (logistic regression).

Results

All DMPs selected for validation showed strong correlations between bisulphite pyrosequencing and Illumina 450k methylation values (r= 0.87-0.97). Global hypomethylation was observed in RT (54.6%) when compared to AM (63.5%, P = 0.021). All probes displayed significant levels of differential methylation in the replication cohort (P = <2.2e-16). No significant associations were observed between DNA methylation and clinicopathological

tumour features, however this may reflect the relatively small number of samples assessed.

Conclusions

Our studies have identified and validated a novel methylomic signature of rectal cancer. Although no clinicopathological correlations were observed with the DMPs investigated, others may represent potential targets in the diagnosis, monitoring and risk stratification of rectal cancer.

Table of Contents

Acknowledgements

First and foremost, I would like to thank Professor Jonathan Mill and Mr Neil Smart for their supervision and guidance throughout the project. Neil, I wouldn't be where I am now without your mentorship and for that I am eternally grateful.

Thank you Dr Kate Ellacott for providing pastoral support and for helping to organise the year. Dr Joe Burrage for his saintly patience and guidance. Dr Eilis Hannon for guiding me through the statistics of the project. Mr Ian Daniels and the HeSRU team for their support both in this project and elsewhere. Dr Rachel Dbeis for laying the foundations of the project. Lidia Romanczuk and Dr Bridget Knight for their support in the of acquisition of tissues, and to Dr Sarah Saunders for providing her histopathological expertise throughout the study. The Exeter Complex Disease group for welcoming me into their team.

Thank you to the Northcott Foundation and Bowel Cancer West for providing the funding required for laboratory consumables and reagents. To the Royal College of Surgeons of England and the Wolfson Foundation for supporting me personally and relieving the financial burden of an additional year of study.

And finally to my friends and family for always being there for me.

List of Figures

6.8	Plots to show the average methylation of all CpGs assessed by each pyrosequencing assay. Red boxes highlight the target DMP/s of each assay. T, tumour; N, non-tumour/ adjacent mucosa.
7.1	Global methylation levels of tumour (T) and non-tumour/ adjacent mucosa (N) tissue samples.
8.1	Histologically confirmed tumour content of historically acquired tissue samples.
8.2	Correlations between mean methylation level of each DMP within the RIST cohort and the historically acquired cohort.
8.3	Scatterplots to illustrate the relationship between tumour load and tumour methylation.
8.4	Scatterplots to illustrate the relationship between tumour load and methylation difference between tumour and adjacent tissue.
8.5	Boxplots to illustrate methylation differences between tumour tissue and adjacent mucosa
8.6	Heatmap to demonstrate methylation values of individual tissue samples at each CpG site. Hierarchical clustering suggests a potential methylomic signature of hindgut (distal) tumours.
8.7	The ability of a logistic regression model to predict tissue type in the RIST cohort
9.1	Boxplots of statistically significant relationships between DMPs and clinicopathological features.
9.2	Ability of logistic regression model (modD) to predict tumour site. Probability values greater than 0.5 correspond to right sided tumours.
10.1	Gel electrophoresis image demonstrating the applicability of pyrosequencing assays to FFPE tissue samples
10.2	Correlation between average methylation values of historical and FFPE tissue samples at Chr1 and Chr8

List of Tables

6.2	Number of samples containing DNA methylation values greater/less than the mean of adjacent mucosa.
6.3	Standard deviation (σ) and coefficient of variance (σ/μ) of the average DNA methylation of individual CpGs within each pyrosequencing assay
8.1	Number of successfully analysed samples and summary statistics of paired t-tests for each DMP.
8.2	Contribution of DMPs to the model when sequentially added.
9.1	Summary of p-values obtained from univariate analysis of relationships between DNA methylation and cliniopathological features. P-values <0.05 are highlighted.

Abbreviations

CpG	Cytosine-guanine dinucleotides
5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
cCR	Complete clinical response
CI	Confidence interval
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CME	Complete mesocolic excision
CMS	Consensus molecular subtypes
CRC	Colorectal cancer
CRF	Clinical Research Facility
CRM	Circumferential resection margin
dH_2O	Deionised water
DMP	Differentially methylated probe
DMR	Differentially methylated region
EMT	Epithelial-mesenchymal transition
EMVI	Extramural vascular invasion
ETB	Exeter Tissue Bank
EWAS	Epigenome-wide association study
FAP	Familial adenomatous polyposis
FFPE	Formalin fixation and paraffin embedding
FOBT	Faecal occult blood test
HNPCC	Hereditary non-polyposis colorectal cancer
HR	Hazard ratio
LCM	Laser capture micro dissection
LINE-1	Long interspersed nucleotide element-1
LUMA	Luminometric methylation assay
modD	Methylation difference (regression model)
modT	Methylation data obtained tumour tissue (regression model)
MRF	Mesorectal fascia
MRI	Magnetic resonance imaging

MSI	Microsatellite instability
nCRT	Neo-adjuvant chemoradiotherapy
NIHR	National Institute for Health Research
OS	Overall survival
PCR	Polymerase chain reaction
pCR	Complete pathological response
RD&E	Royal Devon & Exeter NHS Foundation Trust
RT	Rectal tumour
SNP	Single nucleotide polymorphism
TBE	Tris/Borate/EDTA
TME	Total mesorectal excision
TNM	Tumour node metastasis
TSG	Tumour suppressor gene
TSS	Transcriptional start site
UICC	Union Internationale Contre le Cancer
UTR	Untranslated region

1 Background

1.1 Introduction to epigenetics

1.1.1 Chromatin structure

Each cell contains around two metres of DNA which must be densely packaged into its nucleus. This packaging is achieved through the highly ordered wrapping of DNA around histone proteins to form a structure termed chromatin. The functional unit of chromatin is the nucleosome, which consists of a core histone octamer, a single linking histone and the surrounding DNA (Figure 1.1). These components are held together by electrostatic interactions between the positive histone proteins and negatively charged DNA. Further compaction is permitted through interactions with scaffold proteins.

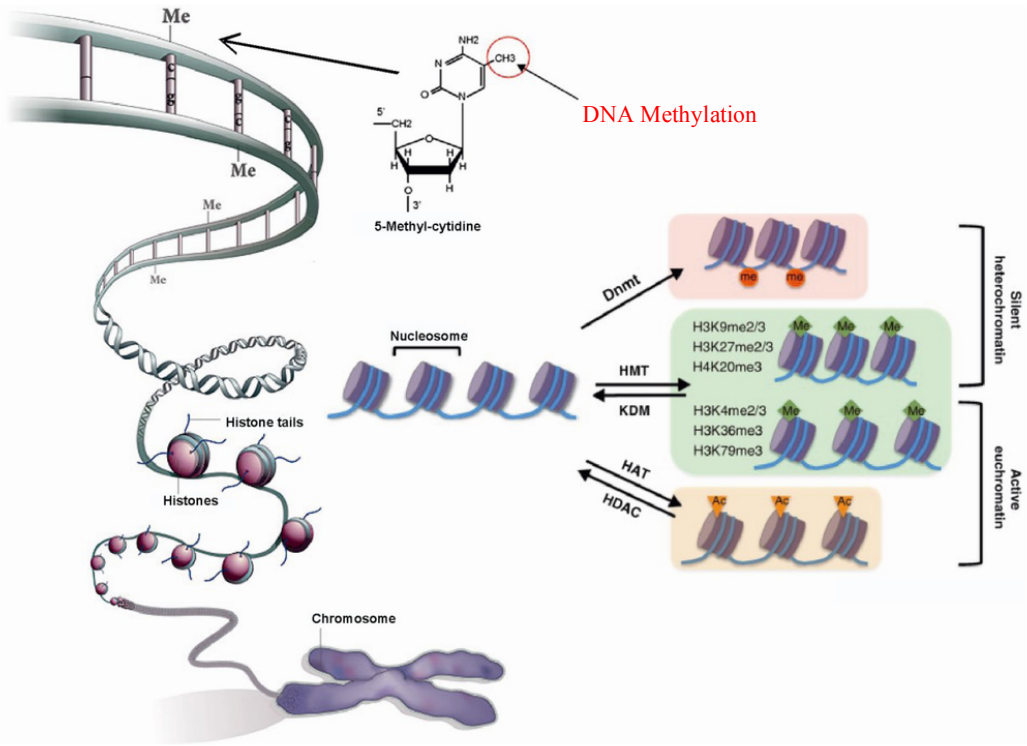


Figure 1.1: Schematic diagram to illustrate chromatin structure. DNA can be seen to wrap around histone proteins to form functional units known as nucleosomes. Key epigenetic modifications, namely DNA methylation and histone tail modifications, are also demonstrated. Taken from Ref (1)

The degree to which chromatin is compressed is modifiable— increasing or reducing the accessibility of DNA to transcription factors. Less densely packed chromatin is termed euchromatin and is associated with gene activation. In contrast, heterochromatin describes densely packed chromatin and is associated with gene silencing. These two states are each influenced by multiple epigenetic modifications.

1.1.2 What is epigenetics?

Literally translated as “upon genetics”, the term epigenetics describes the study of mitotically heritable changes in gene expression, that occur without changes in the underlying genetic sequence.(2) As all cells contain the same genetic code, epigenetic modifications permit the selective expression of cell-type specific proteins and therefore lead to the phenotypic differences observed between genetically identical cells. In addition, epigenetic modifications also contribute to genomic stability and chromosomal structure.

The mechanisms of epigenetic modification are summarised in Table 1.1. Of note, alternative routes to differential protein level and function, such as post-translational modifications, microenvironmental and paracrine effects, also exist.(3) My current study focuses on DNA methylation and therefore other epigenetic modifications are considered beyond the scope of this thesis.

Of note, some variation exists in the relative stability, and tissue specificity, of epigenetic markers.(4) This is an important consideration in the experimental investigation of epigenetics. In a study by Byun et al., the temporal stability of DNA methylation was positively associated with sequence characteristics such as CpG density and 3′ proximity to repeat elements.(5) Tissue specific patterns of aberrant DNA methylation have also been observed in cancer, and are heavily influenced by histone marks in the cell of origin.(6)

Modification	Overview of Mechanism
DNA methylation	Enzymatic addition of methyl group to 5th carbon of cytosine at CpG dinucleotides (see Section 1.1.3). These bring about changes in gene expression through: 1) Interactions with methylated CpG binding proteins to promote the transition from euchromatin to heterochromatin. 2) Directly blocking the binding of transcription factors.
Histone tail modifications (7)	Involve multiple modifications of the N-terminal tails of histone proteins including: Methylation, ubiquitination, phosphorylation, acetylation and sumoylation. These bring about changes in gene expression through: 1) Disruption of the electrostatic charges between DNA and histone proteins. 2) Acting as binding sites for specific proteins.
ATP dependant chromatin remodelling complexes (8)	Use the energy from ATP hydrolysis to move and/or evict histone proteins, resulting in the restructuring of chromatin.
RNA interference pathways (9)	Bring about post transcriptional silencing by the inhibition or degradation of cytoplasmic mRNA by utilising short antisense RNA strands.
Histone variants (10)	Non-canonical variants of histone proteins confer various mechanisms of altered gene expression and structure in a locus specific manner.

Non-coding RNA (11)	Comprise both short (e.g. miRNAs) and long (e.g. lncRNAs) RNAs. They regulate gene expression at the transcriptional and post-transcriptional level.

Table 1.1: An overview of key epigenetic modifications and their mechanisms of bringing about changes in gene expression.

1.1.3 DNA methylation

DNA methylation is the most studied and well understood mechanism of epigenetic modification. It involves the enzymatic addition of a methyl group to the fifth carbon of cytosine to produce 5-methylcytosine (5mC, Figure 1.2).

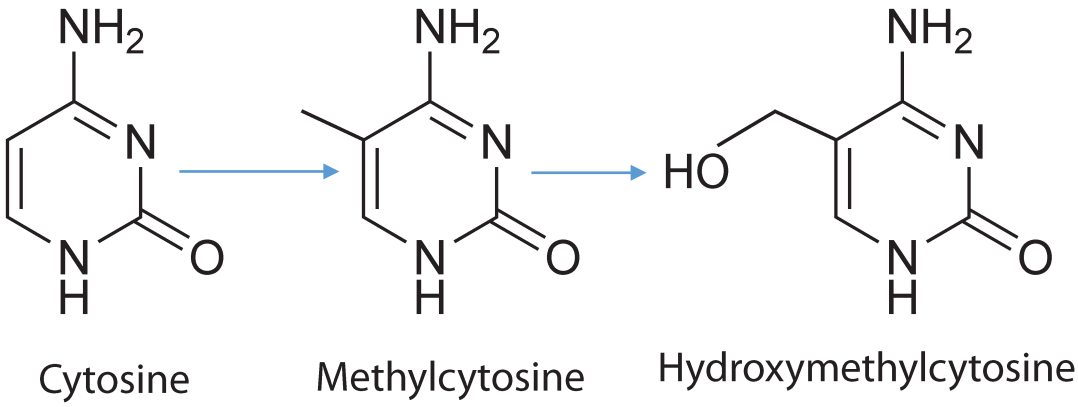


Figure 1.2: Enzymatic addition of a methyl group to the fifth carbon of cytosine by DNA methyltransferase to produce 5-methylcytosine. 5-hydroxymethylcytosine is the first product in 5-methylcytosine demethylation. Importantly, 5-methylcytosine and 5-hydroxymethylcytosine cannot be distinguished by traditional sodium bisulphite treatment and have different effects of gene expression.

DNA methylation occurs almost exclusively at cytosine-guanine dinucleotides (CpGs). This palindromic sequence enables the presence of 5mC on each DNA strand and contributes the mitotic heritability of DNA methylation.

Regions of DNA in which greater than 70% of the sequence is comprised of CpG dinucleotides, are collectively known as CpG islands (CGIs). These occur predominantly at the promoter regions of genes and generally remain unmethylated. When CGIs are methylated, they bring about gene silencing through multiple mechanisms. Interactions with methylated CpG binding proteins, such as *MeCP1* and *MeCP2*, are considered the primary mechanism of promoter silencing. These interactions result in the transition of euchromatin to densely packaged heterochromatin. In some cases, methylation also interferes with gene transcription by directly blocking the

binding of specific transcription factors, however this is considered less significant.(12)

DNA methylation also occurs at intergenic regions and at repetitive elements. In these regions, methylation increases genomic stability through the prevention of transposition and the reduction of transcriptional interference.

The terms hypermethylation and hypomethylation refer to the increased and reduced DNA methylation of CpG dinucleotides respectively.

Of note, demethylation of cytosine occurs through a series of oxidative reactions. The first product in this series, 5-hydroxymethylcytosine (5hmC), is of particular significance in the field of epigenetics as traditional sodium bisulphite treatment (see Section 4.6) cannot distinguish between the two states.(13) In contrast with 5mC, 5mhC is considered to promote gene expression and has been reported to play a role in gene splicing.(3, 13)

1.2 Introduction to Cancer

1.2.1 Cancer terminology

Neoplasia is characterised by the uncontrolled proliferation of abnormal cells and can be benign or malignant in nature. The formation of a malignant neoplasm defines cancer. These neoplasms, or tumours, possess the ability to invade surrounding tissues and metastasise. Metastasis describes the spread of cancer from its primary site to distant organs through the circulatory or lymphatic systems. Although less common, cancer cells can also cross body cavities directly in a phenomenon termed transcoelomic spread.(14) Importantly, benign tumours can progress into malignant tumours with time.

Cancers are broadly classified depending upon their tissue of origin (Table 1.2). Carcinomas account for around 85% of all cancers and originate from epithelial tissue. Adenocarcinoma describes a subclassification of carcinoma originating from glandular epithelium.

Cancer type	Tissue of origin
Carcinoma	Epithelial tissue
Sarcoma	Connective tissue
Leukaemia	Blood cells originating from the bone marrow
Lymphoma	Blood cells originating from the lymphatic system

Table 1.2: A summary of cancer type nomenclature on the basis of tissue of origin.

Hyperplasia represents the initial stage of tumourigenesis and is characterised by an increased proliferation of cells with minimal morphological change. The term dysplasia, describes the presence of change in cellular organisation and nuclear appearance. With disease progression, invasion beyond the basement membrane distinguishes carcinoma in situ from carcinoma.

1.2.2 Cancer as a genetic disease

The pathogenesis of cancer results from the detrimental accumulation of multiple genetic, epigenetic, transcriptomic and metabolomic changes over a period of time. The specific genetic and epigenetic changes observed in colorectal cancer are discussed at length in Section 1.4.

Genetic mutations are the most well understood pathological process of tumorigenesis. Such mutations bring about disease by completely suppressing the expression of, or changing the final structure of, encoded proteins.

With each mitotic division, a cell must fully replicate its DNA and other cellular contents. In normal cells, this process is tightly regulated through a series of cell-cycle checkpoints. At each of these checkpoints, the integrity of DNA is assessed and where abnormalities are identified, the 'decision' to attempt repair or trigger programmed cell death (apoptosis) is made. Key checkpoints between growth phase 1 and synthesis (G1/S), and between growth phase 2 and mitosis (G2/M), each involve multiple proteins involved in the pathogeneis of colorectal cancer such as retinoblastoma protein and tumour suppressor *p53* respectively (see Section 1.4.2).(15, 16) Likewise, the *APC* gene is commonly mutated in colorectal cancer and is particularly important in the pathogenesis of the hereditary condition, familial adenomatous polyposis (see Section 1.4.2).(17) The APC protein is key to the spindle checkpoint, which ensures proper chromatid attachment prior to progression from metaphase to anaphase.(3)

Despite these checkpoints, errors in DNA replication lead to around 10.6×10^{-7} mutations per cell division.(18) This mutation rate is relatively consistent between cell types throughout the body. As a result, Tomasetti et al. describe a strong correlation is seen between the cumulative number of divisions that a cell undergoes over an individual's lifetime, and the risk of developing

cancer.(19)[1] In addition, exposure to environmental carcinogens such as ultraviolet light, cigarette smoke and certain viruses significantly increase the rate at which mutations occur.

Oncogenes and tumour-suppressor genes (TSGs) each promote and inhibit cell growth and proliferation respectively. In health, the expression of these gene types is carefully regulated to bring about controlled cell division during mitosis. In cancer cells however, mutations of these genes result in anomalous activation of oncogenes, and silencing of tumour suppressors. By their nature, loss of function mutations in a TSG can be compensated for by the expression of its functional, complimentary allele.(21) As a result, mutations must occur within both alleles for a cell to become cancerous. This concept was first described by Alfred Knudson and is known as the “two-hit” hypothesis.(3)

Importantly, single gene mutations are insufficient to cause cancer. Instead, multiple mutations of genes involved in several metabolic processes are required. These metabolic processes are collectively termed the ‘Hallmarks of Cancer’— characteristic features initially proposed by Hanahan and Weinberg in 2000, and updated in 2011.(22, 23) The system aims to provide a framework for conceptualising the complex biological characteristics common to all cancer types (Figure 1.3). Through the multistep acquisition of these traits, cells obtain the capacity to proliferate and survive independently—eventually enabling their invasion of local structures and the formation of distant metastases. Attainment of these hallmarks enables the tumour to become more aggressive and allows tumour cells to survive independently.

[1] Of note, the referenced paper by Tomasetti et al. remains highly controversial within the field. The so called “bad-luck” hypothesis argues that 2/3 of human cancers are unavoidable and as a result, public health measures and primary prevention are unlikely to be of value.(19) The study has been criticised however, due to the use of general population incidence data in the United States, without examination of specific at-risk groups such as smokers or alcoholics—both of whom are known to be at greater risk of lung and hepatocellular cancer respectively.(20) The debate holds important clinical relevance as the hypothesis seemingly undermines the importance of lifestyle modification in the prevention of disease.

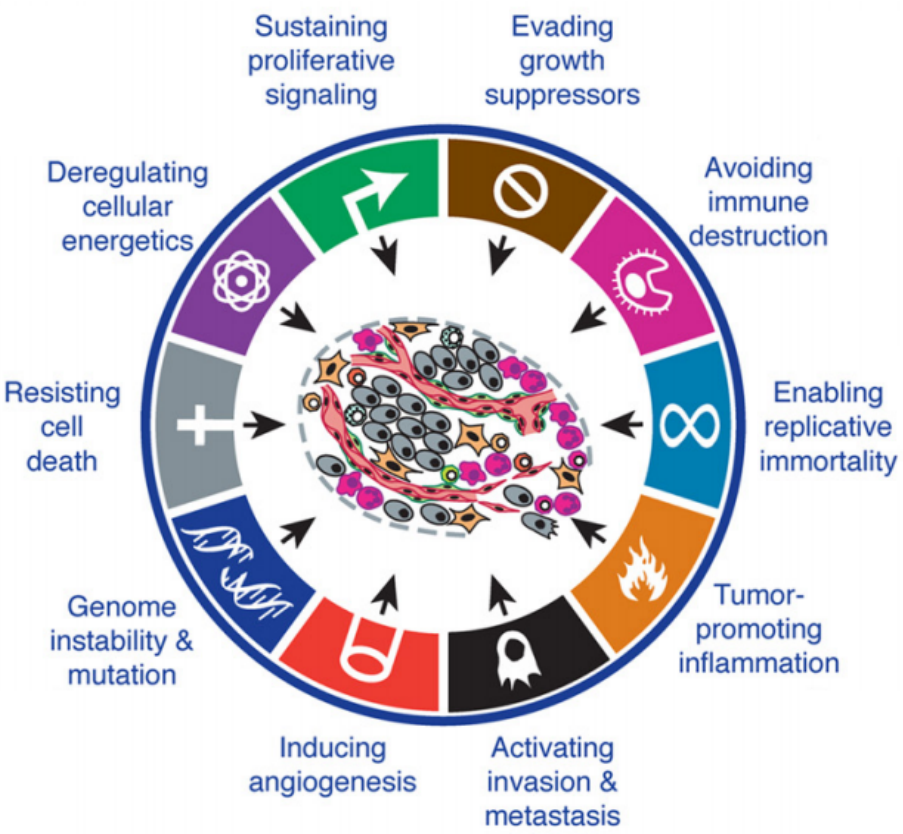


Figure 1.3: A summary of the modified hallmarks of cancer—the common metabolic characteristics that enable cancer cells to survive, proliferate and ultimately metastasise. The hallmarks were initially described by Hanahan and Weinberg in 2000, and updated in 2011.(22, 23) **Figure taken from Ref (23)**

The accumulation of mutations in DNA repair pathways and checkpoints, confounded by the genomic instability possessed by cancer cells, expedites the acquisition of further mutations in subsequent generations of tumour cell. This cascade of increased mutation frequency results in the formation of distinct subpopulations of cells within an individual tumour. According to this so called ‘Big Bang’ model, mutations acquired early in tumourigenesis pervade throughout the tumour whereas those acquired later result in distinct subclones.(3) This phenomenon accounts for the large degree of heterogeneity seen both within and between individual tumours.

Germ-line mutations are present in gametes (sex cells) prior to conception and are inherited by an individual’s offspring. As a result, mutations are passed to all cells during development. In health, germline mutations account

for variation seen between parents and their children. However, the inheritance of certain pathological mutations also predispose individuals to developing cancer. Specific familial colorectal cancer syndromes are discussed briefly in Section 1.4.

1.2.3 Cancer as an epigenetic disease (overview)

In addition to the genetic changes discussed, aberrant changes in the epigenome of patients with cancer can constitute the second hit of Knudson's two-hit hypothesis. This results in abnormal gene expression and genomic instability, which can mimic the phenotypic effects of genetic mutations. Epigenetic changes in colorectal cancer are discussed fully in Section 1.5.

In brief, three main mechanisms underpin the epigenetic basis of cancer as follows:

Global hypomethylation:

Hypomethylation of repetitive sequences, such as long and small interspersed nucleotide elements (LINE and SINE respectively) result in genomic instability.

Hypermethylation of promoter regions:

Hypermethylation of tumour suppressor genes leads to the suppression of their action and thereby promote tumorigenesis. It is now understood that tumour suppressor gene hypermethylation occurs more frequently than genetic mutations.

Direct mutagenicity:

DNA methylation also promotes carcinogenesis directly due to the inherent susceptibility of 5MC to undergo deamination to thymine.(24)

1.3 Colorectal cancer: Background

1.3.1 Anatomy of the colon and rectum

The large intestine refers to the colon and the rectum.(25) The colon begins in the right iliac fossa as the caecum (Figure 1.4). It then ascends superiorly as the ascending colon before turning at the hepatic flexure to form the transverse colon. At the splenic flexure, the colon turns to descend on the left as the descending colon. The sigmoid colon represents the final portion of the colon and is highly variable in length. The sigmoid begins below the level of the pelvic brim where its distal end is relatively fixed at the level of the third sacral vertebrae. The sigmoid colon is continuous with the rectum and as a result, the anatomical distinction between the sigmoid colon and rectum remains a subject of debate.(26) For surgical purposes in the UK, the rectum is arbitrarily defined as the portion of alimentary tract 15cm proximal to the anal verge.(27) Distance from the anal verge is also used to subdivide the rectum into low (0-5cm), middle (5-10cm) and high (10-15cm).(25)

The embryological origin of the colon differs throughout its length.(28) The caecum, ascending colon and proximal two thirds of the transverse colon are derived from the embryological midgut, whereas all distal structures to the level of the proximal anal canal (above the dentate line) are derived from the embryological hindgut. Midgut structures receive their blood supply from branches of the superior mesenteric arteries whereas hindgut structures are supplied by branches of the inferior mesenteric artery. The rectum also receives additional blood supply from the internal iliac arteries via the inferior and middle rectal arteries.(25) Branches subdivide between the muscular layers of the colon before terminal branches enter the mucosa and submucosa. Venous drainage of midgut and hindgut structures is achieved via the portal system via the superior and inferior mesenteric veins respectively.

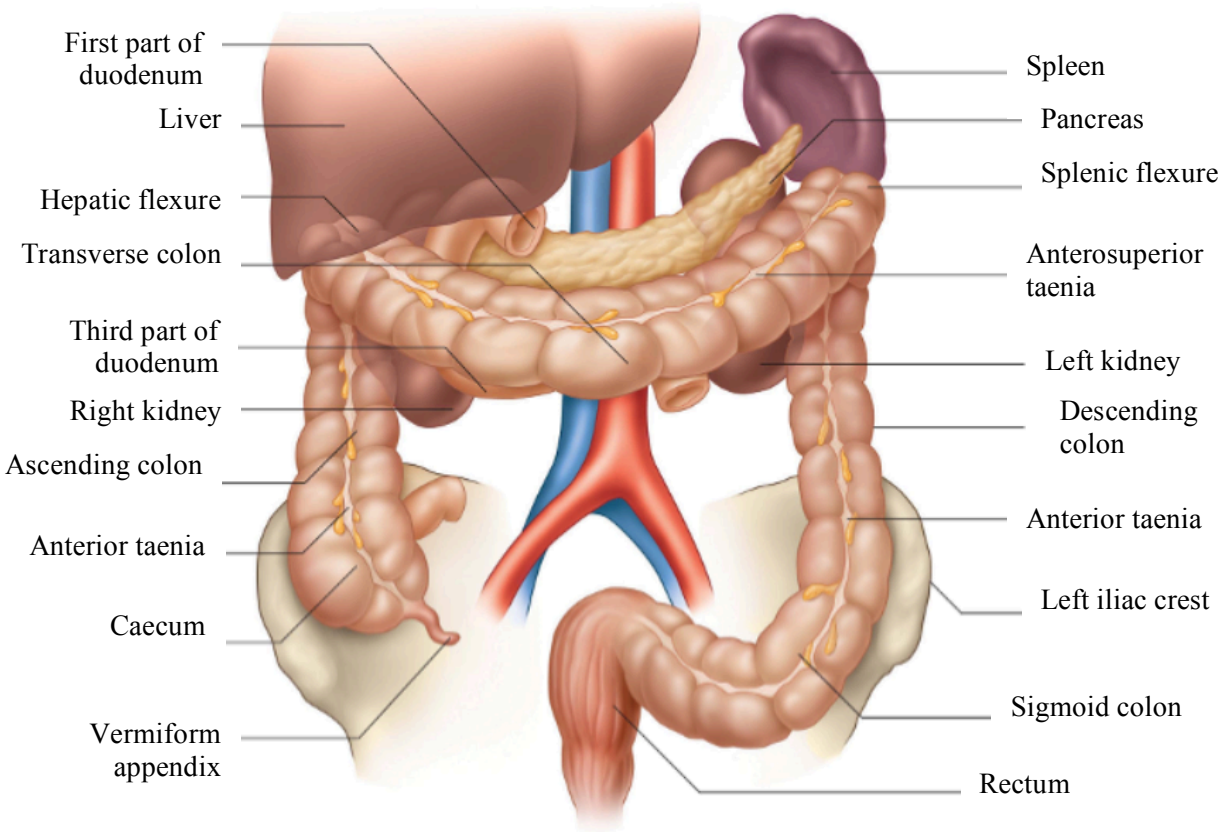


Figure 1.4: Anatomy of large bowel. The large bowel/colon begins in the right iliac fossa as the caecum. It ascends to the hepatic flexure, inferior to the liver where it becomes the transverse colon. At the splenic flexure, the colon descends into the pelvis to become the sigmoid colon, before transitioning into the rectum. The small bowel and mesentery have been removed from this diagram for clarity. Taken from Ref (25)

The colon is surrounded by visceral peritoneum. The ascending and descending colon are fixed within the retroperitoneal cavity. Other areas of the colon are suspended within the peritoneal cavity by mesentery. Until recently, the mesentery was considered a relatively quiescent structure— understood to simply facilitate the passage of neurovascular and lymphatic structures from the retroperitoneum to the bowel. The work of Coffey and O'Leary however, has challenged this traditional notion by demonstrating complex anatomical and physiological features of the structure in health and disease.(29)

The mesentery of the rectum is termed the mesorectum and is of great clinical importance in the staging and management of rectal cancer (Section 1.8). The mesorectum contains the vasculature and lymphatic drainage of the rectum, which act as a gateway to metastatic spread.(30, 31) The mesorectum is encased by an adapted layer of visceral peritoneum termed the mesorectal fascia (MRF), which is separated from the parietal pelvic fascia by a loose layer of areolar connective tissue.(32) This serves as an important landmark in the surgical excision of the rectum.

1.3.2 Histology of the colon and rectum

The wall of the large intestine consists of four layers; the mucosa, submucosa, muscularis externa and serosa.(25) The predominant epithelial cell types in the mucosa are columnar cells and mucinous cells. Columnar cells are responsible for absorption and ion exchange, whereas mucinous cells secrete mucin. The muscularis externa forms an important landmark in the staging of colorectal cancer (Section 1.7).

In parallel to the anatomy of the colorectum, histological differences are also apparent throughout its length. For example, the rectum expresses greater a number of endocrine cells than elsewhere in the large intestine, whereas crypts of the proximal colon are shorter than those found distally.(33) Furthermore, metabolic (e.g. bile acid metabolism) and gene expression (e.g. Na+/H+ antiporter) differences are also well documented.(34)

The potential clinical implications of such differences in the context of colorectal cancer are discussed in Section 1.6.

1.3.3 Incidence of CRC

After lung and breast, colorectal cancer (CRC) is the third most common cancer worldwide (35) and represents the second leading cause of cancer death. In the United Kingdom (UK), 41,265 new cases of CRC were diagnosed in 2014 and 15,903 lost their lives to the disease.(36)

Although significant disparities exist in the recording of cancer incidence between countries (37), the highest incidence of CRC is seen in North America, Western Europe and Australasia where the lifetime risk of developing CRC is around 3-5%.(35) However, the incidence of CRC in previously low risk regions such as eastern Europe and east Asia has increased in recent years—supporting the role of the western lifestyle in tumorigenesis.

In contrast, the incidence of CRC in well developed countries such as the UK and USA has become relatively stable. Between 2003-2014, age standardised incidence of CRC in the UK remained constant for males, but increased by 4% in females.(36) This plateau has been largely attributed to the implementation of screening programmes, with consequent identification and removal of precancerous lesions, but it also reflects changes in population demographics and exposure to risk factors.

1.3.4 Risk factors for developing CRC

Increasing age remains the greatest risk factor in the development of the CRC—representing the accumulation of genetic and epigenetic changes within the bowel and reduced efficacy of DNA repair with time.(3) In the UK, approximately 44% of CRC cases are diagnosed in patients over 75 years old and incidence is highest in patients aged 85-89 years old.(36) As we move towards an aging population, this has important implications for future healthcare provision.

- Males are also at increased risk of CRC—accounting for 55% of cases.(36) Other important risk factors include the presence of inflammatory bowel disease, diabetes mellitus and a positive family history of CRC.(38) Lifestyle factors such as smoking, obesity, excessive alcohol intake and diets rich in red and processed meats also significantly increase risk. Emerging evidence also supports the importance of interaction between host-microbiome and pathogens such as *helicobacter pylori, Streptococcus gallolyticus* and *Fusobacterium spp*, in triggering tumorigenesis.(39)

In contrast, diets high in fibre, exercise and the use of aspirin (a non-steroidal anti-inflammatory drug) and angiotensin converting enzyme inhibitors have each been associated with reduced risk.(40-42)

1.4 The Biology of Colorectal Cancer

1.4.1 Pathogenesis of colorectal cancer

As with other cancers, CRC results from the early development and subsequent accumulation of genetic and epigenetic changes within colorectal epithelial cells.

Three major pathways to CRC have been traditionally described. These pathways are characterised by, and named after, their defining features of chromosomal instability (CIN), microsatellite instability (MSI) and CpG island methylator phenotype (CIMP).(43, 44)

CIN and CIMP pathways generally drive tumourigenesis through the 'adenoma-carcinoma sequence'.(44) These tumours develop in aberrant crypts through the relatively predictable evolution from normal epithelial cells, to polyps, to adenomas and ultimately to adenocarcinomas. However, although the prevalence of adenomatous polyps is around 20%, only around 10% of 1cm adenomas progress to CRC at 10 years.(45) The reasons why some polyps progress and others do not is currently poorly understood.

CIN accounts for around 70-85% of sporadic CRC cases. In traditional descriptions, CIN is characterised by early mutation of the adenomatous polyposis coli (*APC*) tumour suppressor gene, followed by the sequential activation of the oncogene, *KRAS* and inactivation of the tumour suppressor gene, *TP53*.(44) However, it is important to note that this model is in fact an oversimplification of the complex interplay of mutations implicated in the development of the CRC phenotype. In reality, CRC is a highly heterogeneous disease, with individual tumours each possessing around 90 mutations on average.(46) These mutations, in conjunction with defects in DNA damage repair, chromosomal segregation and telomere function result in the defining features of CIN— aneuploidy (imbalanced number of chromosomes) and loss of heterozygosity.

CpG island methylator phenotype, is characterised by the high frequency of aberrant DNA methylation within CpG islands and is discussed fully in Section 1.5.

MSI tumours account for around 15-20% of cases. They result from mutations in genes responsible for DNA repair such as *MSH*2, *MLH*1 and *Exo*1.(43) Their mutation permits the subsequent accumulation of further mutations throughout the genome. These mutations accumulate predominantly in microsatellites— highly repetitive regions comprising one to ten nucleotide tandems found throughout the genome.(3) MSI is more commonly featured in tumours of elderly females and in right sided CRCs. Prognostically, stage II and III MSI tumours are associated with favourable outcomes.(47) Tumours derived from sessile serrated adenomas most commonly display features of MSI. MSI is the predominant pathway responsible for the familial cancer syndrome, hereditary non-polyposis colorectal cancer (HNPCC or Lynch syndrome).

1.4.2 Genetics of CRC

Although most cases of CRC occur sporadically, the disease is associated with a significant heritable component.

Specific inherited CRC syndromes, such as familial adenomatous polyposis (FAP) and HNPCC, have been identified and the specific genetic alterations associated with them have been well described.(48) HNPCC encompasses a spectrum of conditions which result from mutations in one or more DNA mismatch-repair genes, such as *MLH1*, *MSH2*, *MSH6* or *PMS2*.(48) These mutations result in the common phenotype of microsatellite instability. FAP manifests from germline mutations in the APC gene and is characterised by the development of many colorectal adenomas at a young age. Affected patients are at high risk of progression to CRC and usually undergo total colectomy before the age of 30. However, although the risk of developing CRC for individuals affected by these conditions is high, they are collectively responsible for less than five percent of all CRC cases.

Even seemingly sporadic CRCs are often associated with a significant heritable component (see Section 1.4.3). In these cases, positive family history is implicated in around 15-20% of cases. In young patients, the presence of a first degree relative with CRC confers 2-fold risk increase. This risk is increased further for those with more affected relatives and earlier disease onset.(49) In contrast, lifestyle factors such as cigarette smoking and red meat consumption result in only a 6% and 13% increase in relative risk respectively.(50) Clinically, this heritable component is particularly clear in patients with strong family pedigrees for the development of CRC in later life. However, in this group, identifiable genetic conditions are often absent. Such phenotypes may result from yet undiscovered genetic or epigenetic aberrations in one or more genes.

Genes commonly mutated in CRC are summarised in Table 1.3. Those involved in the WNT signalling pathway, such as *APC*, are thought to play a central role in initiating tumorigenesis.(51) Subsequent genetic mutations are often seen to cluster within specific pathways such as MAPK, TGF-B and PI3K-AKT.

Gene or biomarker	Chromosome	Gene Function	Molecular lesion	Frequency (%)
Tumour suppressor genes				
APC	5	Regulates the WNT signalling pathway	Inactivating mutations	40–70
ARID1A	1	Member of the SWI/ SNF family, and regulates chromatin structure and gene transcription	Inactivating mutations	15
CTNNB1	3	Regulates the WNT signalling pathway	Activating mutations	1
DCC	18	Netrin receptor; regulates apoptosis, is deleted but not mutated in colorectal cancer, and its role in primary cancer is still unclear	Deletion or LOH	9 (mutation); 70 (LOH)
FAM123B	X	Involved in the WNT signalling pathway	Inactivating mutations	10
FBXW7	4	Regulates proteasome-mediated protein degradation	Inactivating mutations	20
PTEN	10	Regulates the PI3K–AKT pathway	Inactivating mutations and loss of protein (assessed by immunohisto-chemistry)	10 (mutation); 30 (loss of expression)
RET	10	Regulates the GDNF signalling pathway	Inactivating mutations and aberrant DNA methylation	7 (mutation); 60 (methylation)
SMAD4	18	Regulates the TGFβ _and BMP pathways	Inactivating mutations and deletion	25

TGFBR2	3	Regulates the TGFβ _pathway	Inactivating mutations	20
TP53	17	Regulates the expression of target genes involved in cell cycle progression, DNA repair and apoptosis	Inactivating mutations	50
Proto-oncogenes				
BRAF	7	Involved in the MAPK signalling pathway	V600E-activating mutation	8–28
ERBB2	17	Involved in the EGF–MAPK signalling pathway	Amplification	35
GNAS	20	Regulates G protein signalling	Mutation	20
IGF2	11	Regulates the IGF signalling pathway	Copy number gain and loss of imprinting	7 (mutation); 10 (methylation)
KRAS	12	Regulates intracellular signalling via the MAPK pathway	Activating mutations in codons 12 or 13 but rarely in codons 61, 117 and 146	40
MYC	8	Regulates proliferation and differentiation	Amplification	2 (mutation); 10 (CNV gain)
NRAS	1	Regulates the MAPK pathway	Mutation in codons 12 or 13	2
PIK3CA	3	Regulates the PI3K–AKT pathway	Mutations in the kinase (exon 20) and helical (exon 9) domains	20
RSPO2 and	8 and 6,	Ligands for LGR family receptors, and activate the	Gene fusion and	10

RSPO3	respectively	WNT signalling pathway	translocation	
SOX9	17	Regulates apoptosis	Copy number gain	9 (mutation); <5 (CNV gain)
TCF7L2	10	Regulates the WNT signalling pathway	Gene fusion and translocation	10

Table 1.3: Common genetic mutations in CRC. Adapted from Kuipers et al.(38) *APC*, adenomatous polyposis coli; *ARID1A*, AT-rich interactive domain 1A; *BMP*, bone morphogenetic protein; CNV, copy number variation; *CTNNB1*, catenin β1; *DCC*, *DCC* netrin 1 receptor; *EGF*, epidermal growth factor; *FAM123B*, family with sequence similarity 123B; *FBXW7*, F-box and *WD* repeat domain-containing 7, *E3* ubiquitin protein ligase; GDNF, glial cell-derived neurotrophic factor; GNAS, guanine nucleotide-binding protein, α-stimulating complex locus; *IGF*, insulin-like growth factor; *LGR*, leucine-rich repeat-containing G protein-coupled receptor; LOH, loss of heterozygosity; *MAPK*, mitogen-activated protein kinase; N/A, not applicable; *NDRG4*, *NDRG* family member 4; *PI3K*, phosphatidylinositol 3 kinase; *PIK3CA*, phosphatidylinositol 4,5 bisphosphate 3 kinase catalytic subunit-α; *PTEN*, phosphatase and tensin homologue; *RSPO*, R-spondin; SEPT9, septin 9; *SMAD4*, *SMAD* family member 4; SOX9, SRY (sex-determining region Y) box 9; TCF7L2, transcription factor 7 like 2; *TGFβ*, transforming growth factor-β; *TGFBR2*, TGFβ _receptor 2; *VIM*, vimentin.

1.4.3 Complex genetics of CRC

Importantly, although some forms of CRC can be attributed to single genes, the vast majority of cases occur in a seemingly sporadic nature. These cases are now understood to result from the complex interaction of multiple mutations, with each other and with their environment.(3) Each of these mutations may alone carry only a negligible risk of cancer, however when in combination and in the correct environments, they result in disease.(52) Identification of such loci is challenging due to their individual low risk and relative frequency within a given population. Genome wide association studies (GWAS) have revolutionised our understanding of cancer genetics by

identifying single nucleotide polymorphisms (SNPs) across large patient populations.(3, 53-56) Our understanding of how these loci interact and their potential clinical application however remains in its infancy.(57) A full discussion of the complex genetics of CRC is considered beyond the scope of this thesis, however Figure 1.5 highlights some of the genetic loci implicated to date.(57)

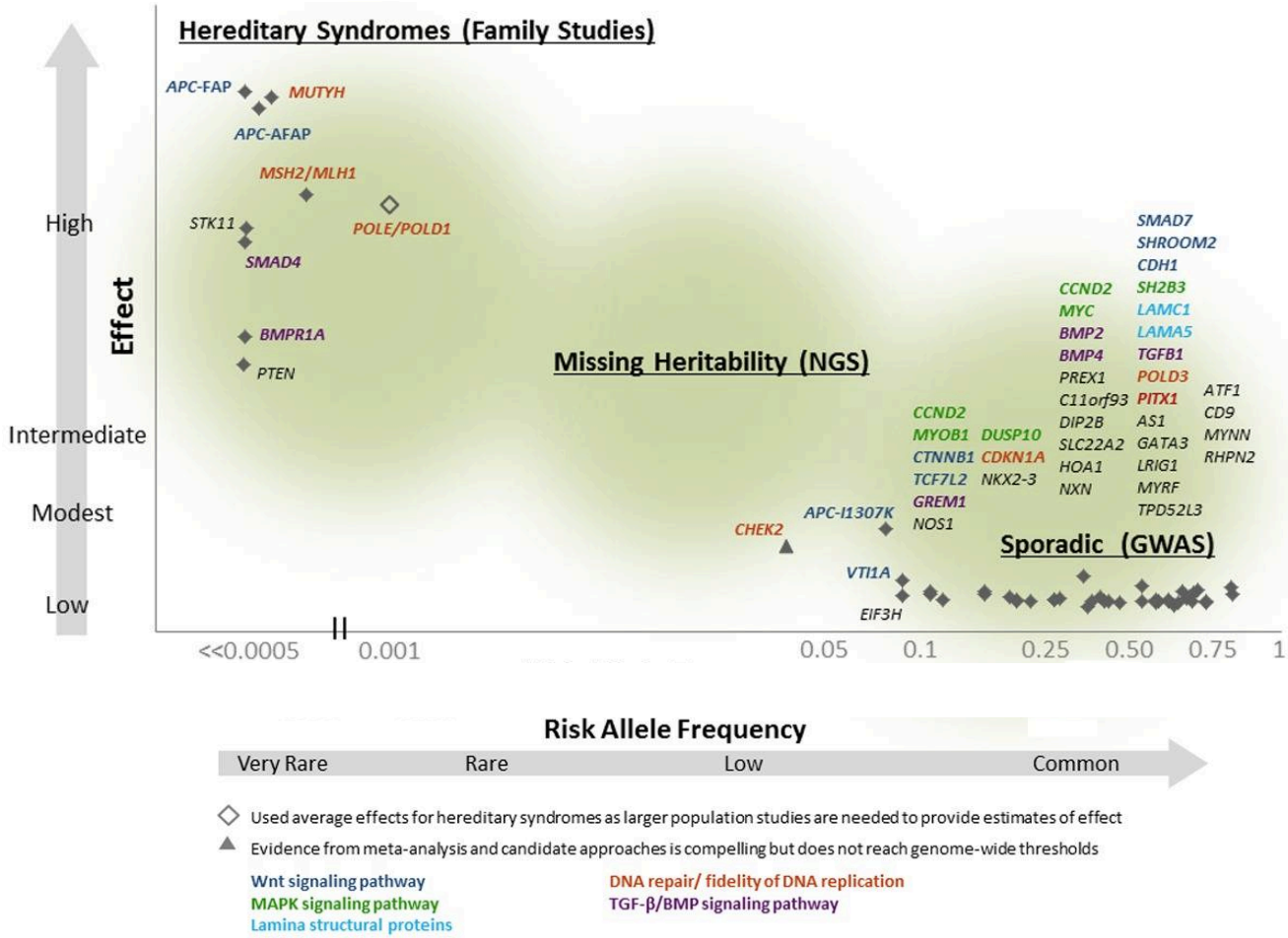


Figure 1.5: Complex genetics of CRC. The tumourigenic loci identified to date range can be broadly classified into two groups. The first comprise rare but high-penetrance mutations. These result in markedly increased risk and are associated with the hereditary CRC syndromes such as familial adenomatous polyposis. The second group consists of common polymorphisms, which each confer weak effects on CRC risk. These are associated with seemingly sporadic CRC. Figure taken from (57)

1.5 Epigenetics of CRC

1.5.1 Hypomethylation

Global reduction of 5-methylcytosine has been widely implicated in carcinogenesis and was initially described in the early 1980s.(58, 59)

DNA hypomethylation is thought to contribute the formation of CRC through multiple pathways. The most well studied mechanism involves the activation of long-interspersed nucleotide element-1 (LINE-1). LINE-1 is a retrotransposon which comprises around 18% of the human genome.(60) The ability of LINE-1 to transpose, brings about chromosomal instability and has been associated with poor outcomes in CRC.(60, 61)

LINE-1 hypomethylation has also been correlated with reduced tumour-lymphocyte infiltration and loss of imprinting which may further contribute to the poor prognosis.(62, 63)

1.5.2 Hypermethylation

As discussed, the epigenetic silencing of tumour-suppressor genes, through the hypermethylation of gene promoters, plays an important role in colorectal tumourigenesis. Multiple studies have demonstrated correlations between specific epigenetic changes and risk factors for CRC development.(64-67) Of note, specific epigenetic markers also appear to affect the susceptibility of patients to develop CRC in response to given risk factors.(68)

Many studies have interrogated the epigenome of CRC to identify potential diagnostic and/or prognostic biomarkers of disease.(69) In the vast majority of these studies, cancers of the colon and rectum are considered together as a single entity. However, as will be discussed in Section 1.6, the biology of CRC differs according to anatomical location of the primary tumour. Despite this, relatively few attempts have been made to assess the epigenome of rectal cancer specifically. These studies are summarised in Table 1.4.

Early studies targeted predetermined genes, based on established understanding of their direct or indirect roles in CRC pathways.(70-74) Commonly investigated genes include *APC, RARB* and *ESR*1.(75, 76) Although significant DNA methylation differences are observed between tumour and adjacent tissue at these loci, their ability to predict clinical outcomes has often failed to reach clinical significance.

More recently, the clustering of multiple genes into panels, has enabled more meaningful conclusions to be drawn. In a study conducted by Gaedcke et al., a high DNA methylation profile (as determined by hierarchical clustering) was associated with increased disease free survival in a cohort of 61 matched rectal tumour and adjacent mucosa tissue samples (HR=3.57, 95% CI= 1.01-12.55, P = 0.0345).(77)

In addition, genome-wide studies have also been performed with the aim of providing a comprehensive methylomic signature of rectal cancer and identifying novel biomarkers for diagnostic and prognostic usage. Multiple studies have utilised the Illumina Infinium HumanMethylation450K assay to meet such goals.(78-82) Of these five studies, only three have attempted to draw correlations with cliniopathologic features (Table 1.4).(79, 80, 82) This method was also utilised in preliminary work conducted by our group (Section 2) with the adoption of more stringent bioinformatic analyses than those reported to date.

1.5.3 CpG island methylator phenotype

CIMP, is characterised by the high frequency of aberrant DNA hypermethylation within CpG islands and is estimated to account for 15-20% of sporadic CRCs. This characteristic is more commonly associated with advanced age, female gender and tumours of the ascending colon.

However, definitions of CIMP vary throughout the literature. In a recent systematic review, 16 individual gene panels were reported to determine CIMP status.(83) As a result, the reported incidence of CIMP also varies

widely from (6.4–48.5 %) although heterogeneous patient demographics and methods of analysis between studies are also likely to contribute to this variation. Commonly cited genes include *CACNA1G, IGF*2, *NEUROG*1, *RUNX*3, *SOCS*1, *CRABP*1, *MLH*1, *p*16, *MINT*1, *MINT*2 and *MINT*31.(83)

Despite this variation, studies have demonstrated prognostic potential, as CIMP-high tumours often exhibit unfavourable outcomes.(84) This is particularly true of rectal cancers, where although relatively rare, its presence has been associated with increased incidence of extramural vascular invasion and metastasis.(85, 86) In addition, interval relapse (defined as relapse within 5 years of screening), microsatellite instability and *BRAF* mutations are also more common in CIMP-high tumours.(87, 88)

1.5.4 Metastasis

Epigenetics also play an important role in tumour metastasis. In order for a carcinoma to metastasise, epithelial cells must acquire the ability to mobilise. Healthy epithelial cells are organised into sheets and are tightly bound to adjacent cells by adherens, desmosomes, tight junctions, and gap junctions.(89) Disruption of these junction molecules and consequent loss of contact with surrounding cells triggers the process of apoptosis.

The epithelial-mesenchymal transition (EMT) describes the process by which mutated epithelial cells dedifferentiate to assume a mesenchymal phenotype.(90) EMT plays an important physiological role in wound healing and embryogenesis, however in cancer, EMT permits the mobilisation and invasion of tumour cells.(91) The ability to redifferentiate into epithelial phenotypes (mesothelial-epithelial transition, MET), however, is critical in the establishment of distant metastases. The reversibility of this process is therefore considered to result from epigenetic modifications.(92, 93) Multiple studies have investigated the effect of specific hypermethylated genes in metastatic rectal cancer. Although some associations have been found, these currently lack the sensitivity and specificity for clinical utility.(94, 95)

Author, Year	Purpose	Technique	Tissues used	Gene selection method	Genes identified	Clinicopathological associations
de Maat, 2008 (96)	Prognostic	Methylation specific PCR	Matched rectal tumour, adenomatous and adjacent rectal mucosa (n=46)	Predetermined	*MINT* panel (*MINT*1,2,3,12,17, 25, 31)	Cluster 3 (node negative, high MINT3, low MINT17) associated with distant recurrence (HR, 2.84; 95% CI, 1.22-6.62; P = 0.02), cancer-specific survival (HR, 3.29; 95% CI, 1.33-8.12; P = 0.01), and OS (HR, 2.21;95%CI, 1.13 to 4.29;P .02) No associations with tumour features or patient demographics
de Maat, 2010 (97)	Prognostic	Methylation specific PCR	Matched rectal tumour, adenomatous and adjacent rectal mucosa (n=251)	Predetermined	*MINT* panel (*MINT*1,2,3,12,17,25, 31)	Combined clusters 1,2,4 associated with local recurrence (HR, 10.23; 95% CI, 1.38-75.91) No associations with tumour features or patient demographics
Leong, 2011(76)	Prognostic	Methylation-specific multiplex ligation-dependent	Rectal adenocarcinoma tissue (n=51), adjacent mucosa (n=35) and	Predetermined	*ESR*1, *CDH*13, *CHFR*, *APC*, *RARB*, *GSTP*1	*RARB* associated with M0 (P = 0·008). *GSTP*1 associated with N0 (P =0.006)

		probe amplification	mucosa of control patients (n=19)			
Molinari, 2013 (75)	Prognostic	Methylation-specific multiplex ligation-dependent probe amplification	Pre-nCRT rectal tumour tissue (n=74), adjacent mucosa (n=16, paired n=9)	Predetermined	*ESR1, CDH13*, *IGSF*4, *APC*, *RARB*, *TIMP*3	*TIMP*3 associated with tumour regression (P = 0.015)
Naumov, 2013 (78)	Diagnostic	Illumina Infinium Human-Methylation 450K BeadChip assay	Matched rectal tumour and adjacent mucosa tissue (n=22), rectal mucosa of control patients (n=19)	Genome wide array	15,667 DMPs (P = <0.05) 171 genes sequentially more methylated from control-> adjacent-> tumour tissue (Δβ >10%)	Ability to distinguish tumour from non-tumour: *ADHFE*1 (multiple CpGs, AUC= 0.990 to 0.998; 95% CI= 0.9771-1), *SND*1 (AUC= 1; 95% CI= 1-1), *OPLAH* (AUC= 0.999; 95% CI= 0.9973-1), *TMEM240* (AUC 0.997; 95% CI= 0.9915-1), *TLX*2 (AUC= 0.974; 95% CI= 0.9546-0.9928), *ZFP*64 (AUC= 0.919; 95% CI= 0.8849-0.9532), *NR*5A2 (AUC= 0.995; 95%

						CI= 0.9872-1), and *COL*4A (AUC= 0.967; 95% CI= 0.9459-0.9878)
Benard, 2014 (95)	Prognostic	Real-time PCR	Discovery cohort: Rectal tumour tissue (n=49), adjacent mucosa (n=10) Validation cohort: rectal tumour tissue (n=88), adjacent mucosa (n=18)	Predetermined	*Apaf*1, *Bcl*2 and *p*53	Increased number of methylated genes associated with reduced OS (HR = 0.28, 95% CI = 0.09-0.83, P = 0.01), cancer-specific survival (HR = 0.13, 95% CI = 0.03-0.67, P = 0.004), distant recurrence-free survival (HR = 0.22, 95% CI= 0.05-0.94, P = 0.001)
Gaedcke, 2014 (77)	Prognostic	MCIp enrichment and CpG island microarray analysis	Discovery cohort: Matched rectal tumour and adjacent mucosa tissue (n=11). Validation cohort: Matched rectal tumour and adjacent mucosa tissue (n=61)	CpG island microarray analysis (Agilent, Germany, Böblingen)	*ADAP*1, *BARHL*2, *CABLES*2, *DOT*1L, *ERAS*, *ESRRG*, *RNF*220, *ST6GALNAC*5, *TAF*4, and *SLC20A*2	High methylation group (determined by heirarchial clustering) associated with increased disease free survival in test cohort (HR=4.09, 95% CI=1.12-14.87, P = 0.0207) and validation cohort (HR=3.57, 95% CI= 1.01-12.55, P = 0.0345)

Kohonen-Corish, 2014 (98)	Prognostic	MethyLight assay	Rectal cancer tissue (n=381)	Predetermined	*CDKN2A*	Multivariate analysis found no association between *CDKN2A* methylation alone and outcomes. *CDKN2A* methylation + *KRAS* mutation associated with reduced OS (HR=2.5, 95% CI= 1.5=4.2, P = <0.001)
Leong, 2014 (94)	Prognostic	Pyrosequencing	Matched rectal tumour and adjacent mucosa tissue (n=133)	Predetermined	*CHFR, CDH*1*, CXCL*12*, APC, MINT*3*, CDH*13*, ESR*1*, UNC5C, GSTP*1, *RARB* and *APC*	*RARB* and *CHFR* associated with T stage (P = <0.001 and 0.005). *RARB, CXCL*12 and *DAPK*1 associated with nodal metastasis (P = 0.008, 0.021 and 0.022). *RARB* associated with LVI (P =0.038) Regression models to predict: LVI (*CDH*1, *CDH13* and *MINT*3) AUC= 0.76 (95% CI= 0.68-0.84); lymph node metastasis (*CDH*1, *CDH*13, *MINT*3, *CXCL*12, *RARB* and *APC*) AUC= 0.76 (95% CI= 0.68-0.84); distant metastasis (*CDH*1, *MINT*3, *CXCL*12, *RARB, ESR*1 and *CHFR*) AUC= 0.82 (95% CI= 0.73-0.91)

Vymetalkova, 2014 (99)	Mechanistic	Methylation-specific PCR, high resolution melting	Matched rectal tumour and adjacent mucosa tissues (n=27)	Predetermined	*MLH1*	Hypermethylation present only in samples with MSI-H (P = <0.001) No association with mRNA expression
Exner, 2015 (100)	Prognostic	Targeted CpG-360 DNA methylation array, methylation-sensitive qPCR	Discovery cohort: Rectal tumour tissue (n=22), matched adjacent mucosa (n=18). Validation cohort: Rectal tumour tissue (n=78), matched adjacent mucosa (n=59).	Targeted CpG-360 DNA methylation array	*TMEFF*2, *PITX*2, *TWIST*1, *ESR*1, *BOLL*, *TFPI*2, *WT*1, *GDNF, HLA-G, PENK, SEZ6L, SFRP*2, *RARB, DCC, GATA*4, *CLIC*4 and *S100A8*	*TFPI*2, *DCC* and *PTGS*2 allowed discrimination between tumour samples and peripheral blood (sensitivity=1, specficity=1, AUC=1). *TMEFF*2, *TWIST*1 and *PITX*2 allowed discrimination between tumour and adjacent tissue (sensitivity= 0.89-1..0, specificity= 0.94-1.0, AUC= 1.0) *CDKN2A* methylation associated with reduced OS (P = 0.017).
Ha, 2015 (79)	Prognostic	Illumina Infinium Human-Methylation 450K BeadChip assay,	Discovery cohort: Post- nCRT rectal tumour tissue (n=45), Validation: Post-nCRT rectal tumour tissue (n=67)	Genome wide array	Discovery: 15 loci associated with TRG1, 25 loci with TRG1-2) and 291 loci with TRG1-3. *DZIP*1, *ZEB*1, *DKK*3, *STL*, *KLHL*34, and	*KLHL*34 associated with response to nCRT (AUC= 0.701, P = 0.036. At 39.7% methylation, sensitivity= 0.625; specificity=0.727)

		pyrosequencing.			*ARHGAP*6 selected for validation.	
Laskar, 2015 (101)	Prognostic	Methylation specific PCR	Rectal tumour tissue (n=80), adjacent rectal mucosa (n=20)	Predetermined	*ASSF*1, *DAPK*, *ECAD*, *BRCA*1, and *GSTP*1.	*RASSF*1 associated with early onset (P = 0.003) and poorly differentiated tumours (P = 0.02), *BRCA*1 associated with late onset (P = 0.02). *GSTP*1 associated with male gender (P = 0.01), early stage tumours (P = 0.04) and late onset (P = 0.001)
Lin, 2015 (80)	Prognostic	Illumina Infinium Human-Methylation 450K BeadChip assay, Sequenom MassCLEAVE	Discovery cohort: Matched rectal tumour and adjacent mucosa tissue (n=23). Validation cohort 1: Matched rectal tumour and adjacent mucosa	Genome wide array	*AGBL*4, *ZNF*625, *MDFI*, *TWIST*1, and *FLI*1	Using panel *of AGBL*4, *FLI*1 and *TWIST*1: Increased number of methylated markers associated with increasing disease stage (P = <0.01) and reduced DFS at 3 years (P = <0.001).

		base-specific cleavage, MassARRAY	tissue (n=75). Validation cohort 2: Blood plasma (n=353)			
Liu, 2015 (102)	Mechanistic	Methylation specific PCR	Matched rectal tumour and adjacent mucosa tissue (n=40)	Predetermined	*CHD5*	*CHD5* methylation associated with reduced gene expression (P = <0.05)
Vymetalkova, 2016 (81)	Mechanistic	Illumina Infinium Human-Methylation 450K BeadChip assay, pyrosequencing	Matched rectal tumour and adjacent mucosa tissue (n=25)	Genome wide array	5929 DMPs identified (P = <0.05). *TIFPI2*, *HBBP1*, *ADHFE1*, *BPIL3*, *FLI1* and *TLX1* validated	None reported
Wei, 2016 (82)	Diagnostic	Illumina Infinium Human-Methylation 450K BeadChip	Discovery cohort: Matched rectal tumour and adjacent mucosa tissue (n=6). Validation cohort:	Genome wide array	18,568 DMPs identified (P = <0.05) Genes with >2 DMPs within promoter: *EYA4*,	*GFRA1* and *GSTM2* not associated with stage. Ability to distinguish tumour and adjacent tissue: *GFRA1* (AUC= 0.949), *GSTM2* (AUC= 0.926)

		assay, methylation sensitive high resolution melting.	Matched rectal tumour and adjacent mucosa tissue (n=44)		*GFRA*1, *FOXI*2, *SLITRK*1, *STOX*2, *CNRIP*1, *SFRP*1, *ADHFE*1, *C2orf*40, *KCNC*2, *KCNQ*1, *LONRF*2, *MEST*, *RALYL*, *HKDC*1, *KCNIP*4 , *SORCS*1, *CBLN*2, *FRZB*, *GALR*1, *PMEPA*1, *RARRES*2, *SLC6A*5, *AZGP*1, C10orf81, *FAM*110A, *GLRA*3, *GSTM*2, *HSD11B*1, *MAL*, *PHACTR*3, *SST*, *TMEFF*2, *TNFRSF*8, *TUSC*3, *ZNF*655	
Hua, 2017 (103)	Mechanistic	Bioinformatic	miRNA, mRNA and DNA methylation data obtained from the	Inverse relationship between methylation	*SORCS*1, *PDZRN*4, *LONRF*2, *CNGA3*, *HAND*2, *RSPO*2 and *GNAO*1	None reported

			Cancer Genome Atlas (n=155)	and expression		
Yokoi, 2017 (104)	Prognostic	Quantitative methylation specific PCR	Pre-nCRT rectal tumour tissues (n=33), Matched post-nCRT adjacent mucosa (n=33)	Based on gene expression studies of radiation sensitive and resistant CRC cell lines	*CRBP*1, *STC*2 and *SLCO*3A1	*CRBP*1 methylation associated with tumour regression (P = 0.031)

Table 1.4: Summary of the available literature investigating differentially methylated genes in rectal cancer. Studies investigating the role of CIMP exclusively have been excluded. n, number of patients; TRG, Tumour regression grade; nCRT, neoadjuvant chemoradiotherapy; PCR, polymerase chain reaction; AUC, area under receiver operated curve; OS, overall survival; LVI, lymphovascular invasion; EMVI, extramural vascular invasion; HR, hazard ratio; DFS, disease free survival.

1.6 The Impact of Tumour Location in CRC

1.6.1 Colon and rectal cancers as distinct clinical entities

The rectum is the most common site of CRC, accounting for around 28% of cases.(105, 106) However, it has been largely accepted that colon and rectal cancers are discrete clinical entities with differing aetiology, prognosis and management. As discussed in Section 1.3, multiple anatomical and histological differences exist between the colon and rectum. In addition, the local conditions to which the mucosal epithelial cells are exposed, also vary according to site.

Epidemiologically, lifestyle factors appear to have a greater oncogenic impact on the colon than the rectum.(107) Vast differences also exist in the management of colon and rectal cancer, predominantly as a result of anatomical restraints of the rectum within the narrow pelvis.(108) The management options of colon and rectal cancers are discussed further in Section 1.8.

With regard to underlying tumour biology however, distinct dissimilarities between colon and rectal cancers are less apparent, with tumours of the rectum and distal colon often sharing characteristics. In contrast, clear biological distinctions have been demonstrated between proximal and distal CRCs as discussed below (Section 1.6.2).(109, 110)

The Cancer Genome Atlas Network performed genome-wide analysis of mutations present in 224 matched tumour/adjacent mucosa tissue pairs.(110) In this study no clear differences were observed between tumours of the rectum and distal colon. These findings mirror those of other studies which support the notion of common mutational pathways between distal colon and rectal tumours.(109)

1.6.2 Site dependent variation in features of colon cancers

The concept of pathophysiologically distinct subgroups of CRC cancer, on the basis of anatomical location, was initially proposed by Bulfill in 1990.(111)

Bulfill suggested that tumours proximal and distal to the splenic flexure possessed discrete molecular features. Although some recent studies have suggested a more subtle transition, this dichotomous divide remains the most commonly described pattern.(112, 113)

As discussed in Section 1.3, the proximal and distal colon are derived from the embryological midgut and hidgut respectively. During development, these embryological differences confer different patterns of gene expression, which may underpin observed differences in the underlying tumour biology of proximal and distal CRCs in adult life.(114)

In addition, multiple environmental differences throughout the length of the bowel may also influence underlying tumour development and biology. Although the microbiome of any individual patient is relatively consistent throughout the colon, Flemer et al. recently reported significant differences in the microbiome of patients with proximal (increased abundance of Faecalibacterium, Blautia and Clostridium) and distal CRCs (increased Alistipes, Akkermansia, Halomonas and Shewanella).(115) Although in its infancy, the ability of the gut microbiome to influence DNA methylation, histone modifications and non-coding RNAs has become increasingly recognised in recent years.(3, 116) Exposure to other mutagenic substances such as deoxycholic acid (a secondary bile salt), also differ between sides.(117, 118) As epigenetic modifications can be influenced by environmental factors, the epigenetic signature of CRCs may also vary according to location.

Contrasting exposure to environmental factors and functional requirements necessitates differential gene expression throughout the colon, which is apparent even in healthy individuals.(119) Similarly, specific genetic and epigenetic profiles are also apparent between proximal and distal tumours. Whereas point mutations in *APC*, *NRAS* and *TP*53 are more common in distal tumours, those of *BRAF*, *TGFBR*2, *FBXW7* and *PIK3CA* are more common in proximal disease.(110, 120) Other, more general genetic features are summarised in Table 1.5.

Tumour Site	CIMP-High	MSI-High	MLH1 Hypermethylation	CIN
Proximal	High	High	High	Low
Distal	Low	Low	Low	High

Table 1.5: The Prevalence of key genetic and epigenetic features in proximal and distal CRCs. Adapted from Lee et al. (114) CIMP, CpG island methylator phenotype; MSI, microsatellite instability; CIN, chromosomal instability.

Multiple studies have sought to assess the prognostic value of left and right sided diagnoses. Weiss et al. reported a stage-specific pattern whereby proximal stage II tumours, but distal stage III tumours, demonstrate favourable outcomes.(121)

In metastatic disease, Zhang et al. recently reported that palliative resection of left sided tumours prolonged overall survival by eight months (P = 0.009), whereas no benefit was attained from resection of right sided tumours (P = 0.91).(122) In a meta-analysis of 2977 cases, Cao et al. reported increased overall survival of patients with left-sided, wild type *KRAS* tumours, treated with cetuximab in comparison to patients with right sided tumours.(123) Additional studies have also demonstrated location dependant sensitivity to specific chemotherapy regimens, which largely results from the site-dominant prevalence of specific mutations.(124) Key differences between proximal and distal CRCs are summarised in Figure 1.6.

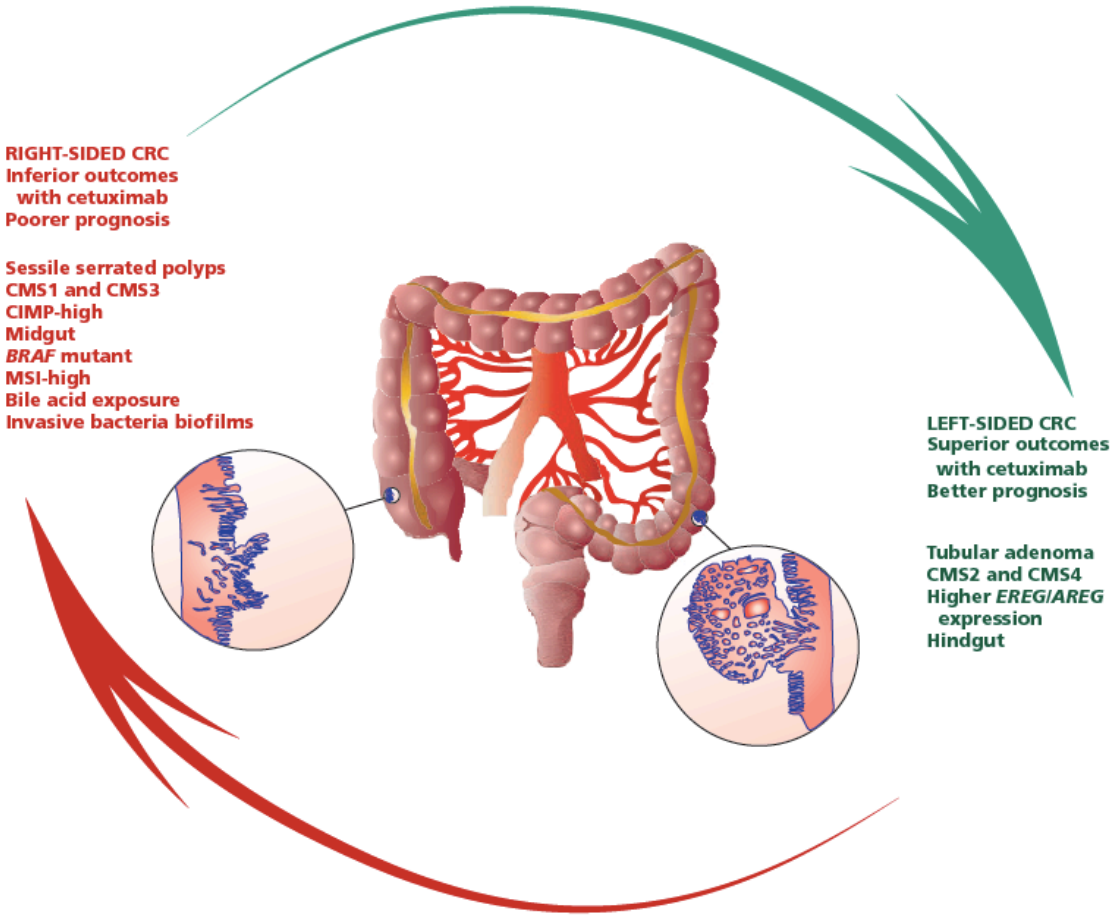


Figure 1.6: Summary of key differences in clinicopathological tumour characteristics between proximal and distal CRCs. Tissues of the proximal and distal colon have different embryological origins, namely the midgut and hindgut respectively. Further differences exist in the local environment and microbiotia of the proximal and distal colon, which may further contribute to observed differences in underlying tumour biology. Taken from Ref (114)

1.7 Colorectal cancer screening and diagnosis

1.7.1 Screening for CRC

The protracted time taken for CRCs to develop creates a window of opportunity for early diagnosis. Early identification and subsequent removal of pre-cancerous lesions has resulted in reduced disease specific mortality in developed countries.(125, 126)

Of note, it is important to acknowledge the concept of “lead time bias” in the context of screening. This describes the process by which early identification of disease gives the artificial impression of prolonged survival, without significant modification of the disease course.(127) This phenomenon is often neglected from studies and should therefore prompt cautious interpretation in improved survival outcomes.

Effective screening techniques must be sensitive, safe, relatively inexpensive and considered acceptable by patients.(45) The faecal occult blood test (FOBT), sigmoidoscopy and colonoscopy are the most commonly cited qualifiers of these criteria and each allow for earlier identification of CRC in average-risk populations. (126, 128) In 2015, various combinations of these tests had been adopted into national screening programmes of 24 of 28 European Union nations.(129)

1.7.2 Epigenetic biomarkers in colorectal cancer

CRC remains largely asymptomatic in its early stages. As a result, 52-56% of cases in the UK are diagnosed at stages III or IV, and 23-26% have established metastases by the time of diagnosis.(36)

Molecular biomarkers offer a non-invasive, easily obtainable alternative to traditional screening methods. The clinical utility of potential biomarkers obtained from blood, stool and urine have each been investigated with variable success.(130) Multiple potential epigenetic biomarkers have been identified. Circulating *SEPT*9 promoter hypermethylation is the most well-reported epigenetic biomarker and is available for use in clinical practice.(131)

However, although early studies described sensitivity and specificity of 72% and 90% respectively, subsequent studies have failed to reproduce such favourable statistics.

Heiss et al. performed genome-wide methylation profiling of the DNA extracted from leukocytes of 139 patients with colorectal cancer and 140 controls across screening and clinic populations.(132) They identified significant differential DNA methylation in the promoter region of the *KIAA1549L* gene. Logistic regression models calculated discrimination (as measured by c-statistic) as 0.69 and 0.73 between CRC cases in the screening setting and controls, and CRC cases in the clinical setting and controls, respectively. Although this would be insufficient to qualify for screening of patients, the study demonstrates the feasibility of such endeavours.

Pellise et al. demonstrated the feasibility of hypermethylated gene promoter analysis to detect occult neoplastic cells in fine needle aspirate, however the genes analysed in this study (*GMT, INK4a and ARF*)[2] failed to match the sensitivity and specificity of conventional cytology.(134)

Greater understanding of the biological and clinical implications of specific epigenetic aberrations in CRC may enable the delivery of personalised care.

1.7.3 Diagnosis

Colonoscopy remains the gold standard for reasons described above.(135) Other methods including capsule endoscopy (sensitivity 88%, specificity 82%) and CT colonography (sensitivity 96%) each carry additional risks and benefits and are therefore usually reserved for patients unable/unwilling to undergo endoscopy.

[2] Of note, ARF and p16 comprise alternatively spliced forms of the gene *CDKN2A*.(133)

1.7.4 Classification and staging of colorectal cancer

CRCs are classified according to the TNM system (Table 1.6). These classifications are then combined into an overall Union Internationale Contre le Cancer (UICC) stage (Table 1.7), which provides valuable prognostic information and is used to guide management. The modified Dukes' classification system is based on the anatomical spread of a tumour and is summarised in Table 1.8.

On an epidemiological scale, increasing TMN and/or modified Dukes' stage each confer worse prognoses.(136) On an individual basis however, their prognostic value is limited. As such, recent years have witnessed an increased interest in determining more educated systems of classification based on biological characteristics.

Primary tumour (T)	
Primary tumor cannot be assessed	TX
No evidence of primary tumor	T0
Carcinoma in situ: intraepithelial or invasion of lamina propria	Tis
Tumor invades submucosa	T1
Tumor invades muscularis propria	T2
Tumor invades through the muscularis propria into the pericolorectal tissues	T3
Tumor penetrates to the surface of the visceral peritoneum	T4a
Tumor directly invades or is adherent to other organs or structures	T4b
Regional Lymph Nodes (N)	
Regional lymph nodes cannot be assessed	NX
No regional lymph node metastasis	N0
Metastasis in 1-3 regional lymph nodes	N1
Metastasis in one regional lymph node	N1a
Metastasis in 2-3 regional lymph nodes	N1b
Tumour deposit(s) in the subserosa, mesentery, or nonperitonealised pericolic or perirectal tissues without regional nodal metastasis	N1c
Metastasis in four or more regional lymph nodes	N2
Metastasis in 4-6 regional lymph nodes	N2a
Metastasis in seven or more regional lymph nodes	N2b
Distant Metastasis (M)	
No distant metastasis	M0
Distant metastasis	M1
Metastasis confined to one organ or site (eg, liver, lung, ovary, nonregional node)	M1a
Metastases in more than one organ/site or the peritoneum	M1b

Table 1.6: Tumour node metastasis (TNM) classification of colorectal cancer. The classification is based on the anatomical spread of the tumour, and can be made histologically or radiologically. This information is then used to calculate a Union Internationale Contre le Cancer (UICC) stage which can inform treatment (Table 1.7).(137)

Stage	T	N	M
0	Tis	N0	M0
I	T1	N0	M0
	T2	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
IIIA	T1-T2	N1/N1c	M0
	T1	N2a	M0
IIIB	T3-T4a	N1/N1c	M0
	T2-T3	N2a	M0
	T1-T2	N2b	M0
IIIC	T4a	N2a	M0
	T3-T4a	N2b	M0
	T4b	N1-N2	M0
IVA	Any T	Any N	M1a
IVB	Any T	Any N	M1b

Table 1.7. Union Internationale Contre le Cancer (UICC) colorectal cancer staging system. The UICC is based upon the tumour node metastasis system and provides valuable prognostic information and can be used to guide management. Taken from Ref (137)

Modified Dukes' Stage	Features	5-year relative survival (%)	Confidence interval (95%)
A	Limited to muscularis propria	93.2	92.5 - 93.9
B	Extending beyond muscularis propria	77	76.4 - 77.5
C	Lymph node involvement	47.7	47.1 - 48.3
D	Distant metastatic spread	6.6	6.1 - 7.0

Table 1.8: Modified Dukes' classification system. The system is based on the anatomical spread of the cancer; whereby more advanced stage confers poorer outcomes. Although the modified Duke's classification is still commonly used in clinical practice, the TNM system is more informative and has largely replaced the Duke's system with regard to informing management. Survival outcomes obtained from Ref (138)

1.7.5 Molecular classification of colorectal cancer

Our greater understanding of the underlying pathophysiology, and the genetic and epigenetic changes seen in cancer; coupled with our increased ambition to determine their clinical significance, has fuelled the field of personalised oncology. In CRC, the search for consistent and reproducible subclassifications has been hindered in part by the large degree of clonal heterogeneity seen both within and between individual tumours.(139, 140) However, progress has been made.

In 2015, the International CRC Subtyping Consortium (ICRCSC) assessed interconnectivity between of six independently reported subclassification systems.(141) This work resulted in the determination of four 'consensus molecular subtypes' (CMS) of CRC, the features of which are summarised in Table 1.9.

The system has demonstrated prognostic value, with unfavourable outcomes associated with the CMS 1 and 4 phenotypes. However, although this classification system holds promise, it has yet to become widely adopted into clinical practice. The impact of stromal contamination on classification has

drawn criticism from some, although the influence of stromal gene expression signals on sample classification is considered to be low.(142) In addition, the system depends upon multiple methods of analysis including histopathological, genetic and epigenetic analysis. Such analyses are inherently costly and time consuming—factors which again limit their clinical utility. Simplification and streamlining of subclassification systems are therefore desirable.

To our knowledge, no rectal cancer-specific molecular subclassification systems have been described to date. Greater understanding of the molecular, epigenetic and transcriptomic processes involved in rectal cancer, and their impact on clinical outcomes and treatment responses, may facilitate the determination of such systems.

CMS	Proportion/%	Features
1 (MSI immune)	13	Microsatellite instability, hypermutated genome, high frequency of BRAF mutations, high immune infiltrate. More common in females, right sided tumours, high histopathological grade. Poor prognosis.
2 (Canonical)	37	Somatic copy number aberrations, HNF4A enrichment and upregulation of WNT and MYC pathways. Mainly left sided tumours.
3 (Metabolic)	13	Disruption of metabolic pathways, chromosomal instability, few somatic copy number aberrations, hypermutation, CIMP-low, high frequency of KRAS mutations.
4 (Mesenchymal)	23	Upregulation of genes involved in epithelial-to-mesenchymal transition, angiogenesis, matrix remodelling and compliment activation. Late diagnosis and high recurrence.

Table 1.9: Consensus molecular subtypes (CMS) of colorectal cancer and histological, epi/genetic and clinical features of each. Of note, these do not add up to 100% as a result of the difficulty in classifying some tumours. MSI, microsatellite instability.

2 Preliminary results

Preliminary experiments were performed by Dr Rachel Dbeis at the University of Exeter Medical School during the 2015-2016 academic year.

Matched rectal tumour (RT) and adjacent normal mucosa samples were prospectively obtained from 15 patients (RIST cohort, Section 4.2) and flash frozen prior to storage. DNA was treated with sodium bisulphite and analysed using the Illumina Infinium HumanMethylation 450k Beadchip array. Additional samples were obtained from a historically acquired cohort and preserved by formalin fixation and paraffin embedding (FFPE) prior to DNA extraction. However, this DNA failed to meet the required quality for genome wide analysis.

DNA methylation data from the RIST cohort underwent stringent quality control using novel bioinformatics approaches. A linear regression model was used to identify differentially methylated probes (DMPs) and genomic differentially methylated regions (DMRs) associated with rectal cancer.

Global DNA methylation levels of RT were decreased in comparison to adjacent mucosa. 176 highly statistically significant DMPs were identified ($p < 1E-07$, Figures 2.1 and 2.2, Appendix 1). 79% of DMPs were located in CpG islands associated with gene promoters and silencing of regulatory genes. Region level analysis was performed using Comb-p to identify spatially correlated regions of differential DNA significantly associated with rectal cancer ($P < 1E-07$, number of probes $>=5$).(143) In total, 828 DMRs were identified between rectal cancer and normal tissue.

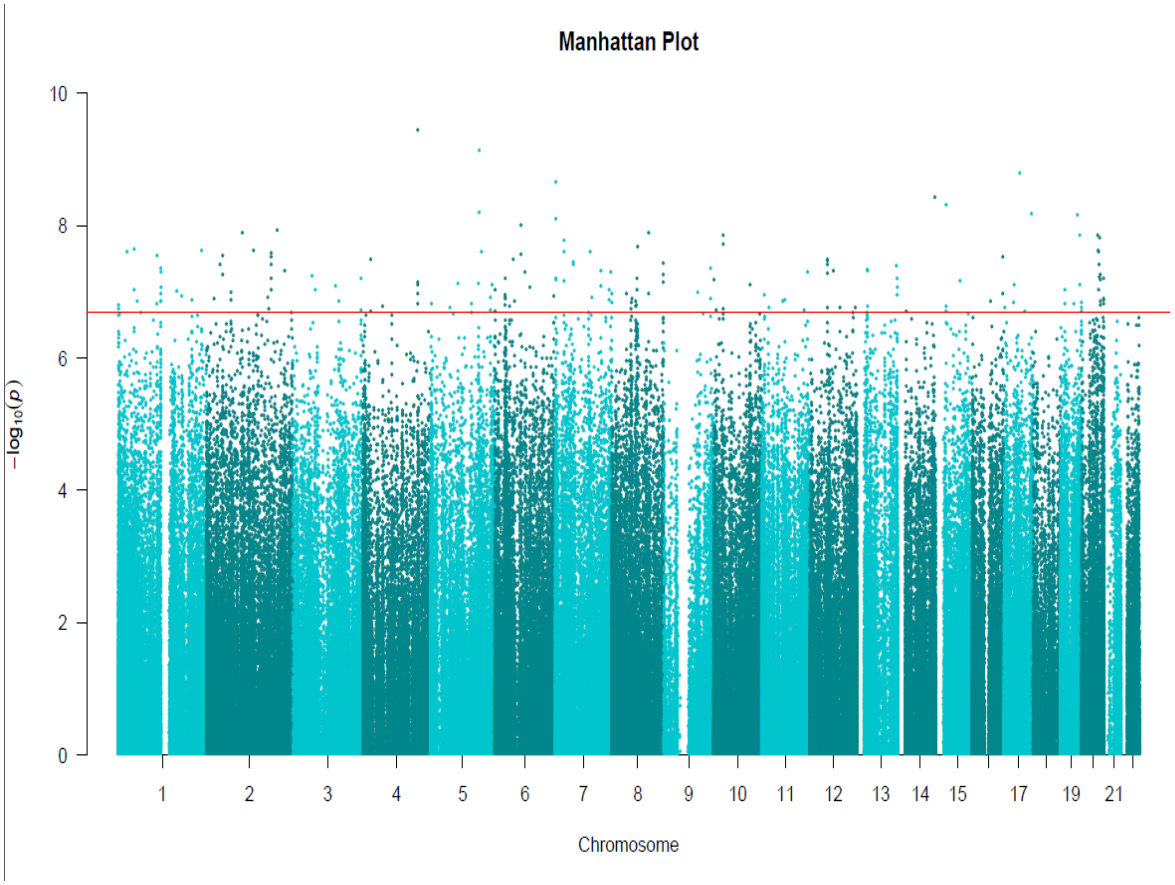
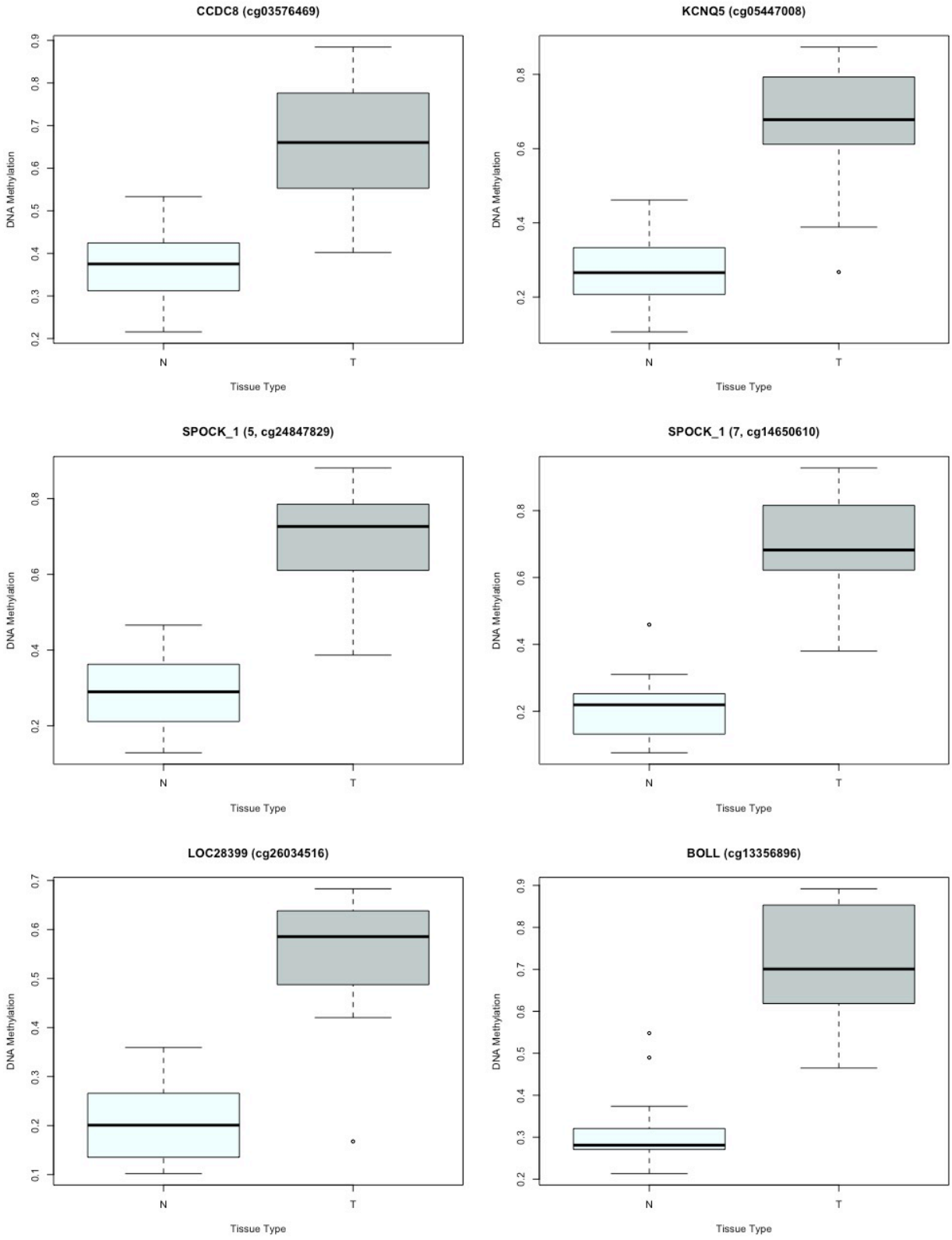


Figure 2.1: Manhattan plot showing the distribution of tumour associated DMPs across all autosomes. The red line represents the threshold (P < 1.25E-7). The dots above the red line portray the DMPs and the dots below the line depict all probes analysed on each chromosome.

Methylation Differences of DMPs Explored in the Current Study as Determined by Illumina 450k Methylation Array



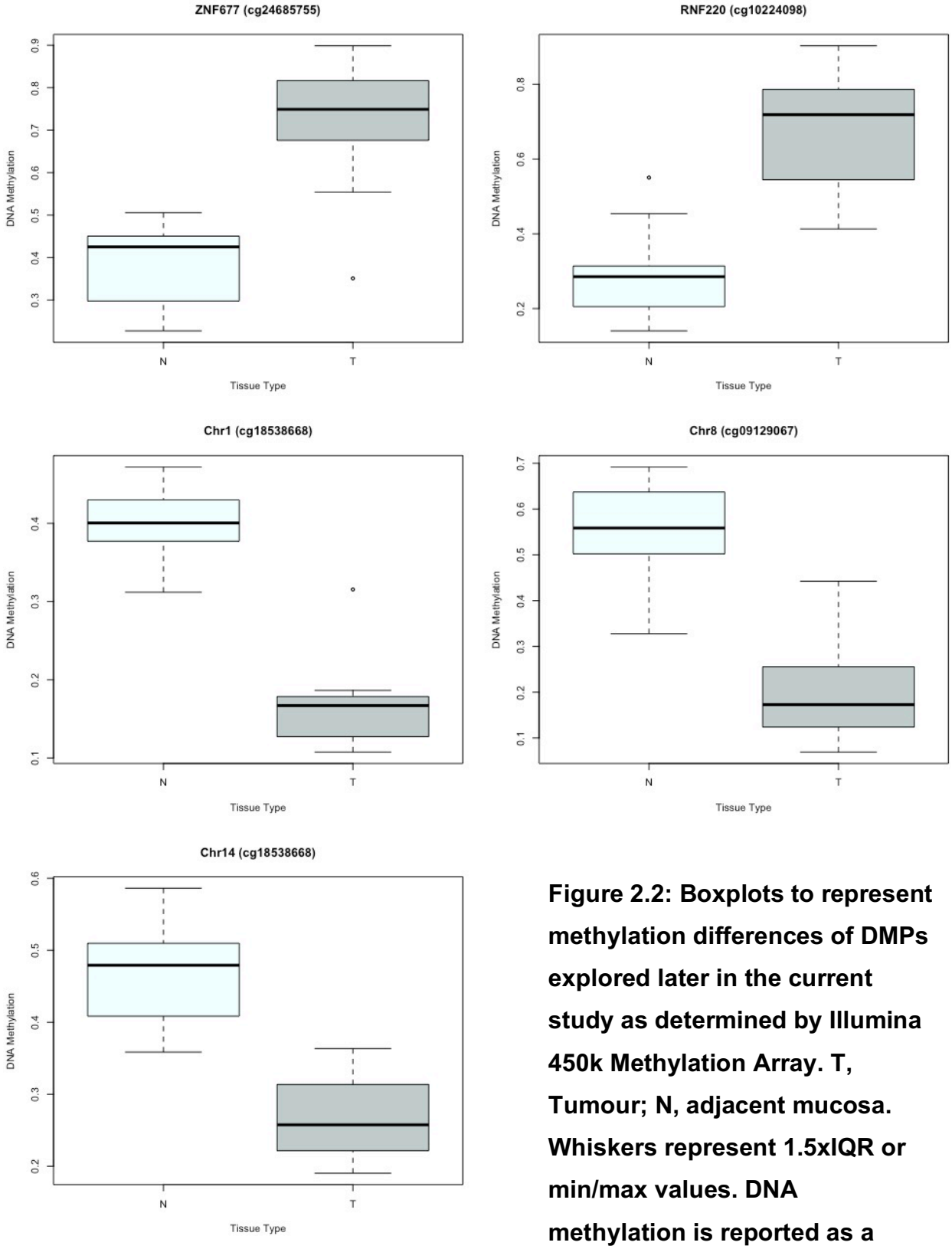


Figure 2.2: Boxplots to represent methylation differences of DMPs explored later in the current study as determined by Illumina 450k Methylation Array. T, Tumour; N, adjacent mucosa. Whiskers represent 1.5xIQR or min/max values. DNA methylation is reported as a proportion of methylated CpG sites / total CpG sites analysed.

3 Aims and objectives

3.1 Purpose of this research

Rectal cancer is a highly heterogeneous disease with respect to underlying tumour biology and to clinical outcomes. Although tumour features such as Dukes' stage offer some prognostic value on an epidemiological level, our ability to predict tumour behaviour and clinical outcomes for individual patients is limited. This work primarily aims to assess the feasibility of a larger study to investigate the relationship between DNA methylation and response to neoadjuvant chemoradiotherapy (see Future work, Section 11.7).

To date only a limited number of studies have conducted genome wide DNA methylation analysis with the aim of producing a rectal cancer specific methylomic signature. Our preliminary work is unique in its use of stringent bioinformatic tools to conduct its analysis. The first phase of this project was to validate the findings of our preliminary work through the development and application of specific bisulphite pyrosequencing assays.

Subsequent replication in a larger cohort of mixed colon and rectal tumour samples was intended to support the biological reproducibility of our preliminary work. This stage also intended to determine the specificity of DMPs identified to rectal cancer, and to provide sufficient sample sizes for the assessment of the relationship between clinicopathological features with methylomic changes seen.

Greater understanding of the underlying biology of rectal cancer and its clinical implications, may enable the development of improved screening, diagnostic and prognostic biomarkers for use in clinical practice.

3.2 Research questions

1. Do the DNA methylation differences identified at specific CpG sites in our preliminary epigenome-wide association study (EWAS) reflect true biological differences between tumour and adjacent tissue?

2. Do the global DNA methylation differences observed in our EWAS reflect true biological differences between tumour and adjacent tissue?
3. Do the DMPs identified in our EWAS exist between other colorectal tumour and adjacent tissue samples?
4. Are the DMPs identified in our EWAS associated with clinicopathological tumour features?

3.3 Specific aims and objectives

The work outlined in this thesis comprises four key components (Figure 3.1). These are as follows:

Phase 1: To validate DNA methylation differences identified by preliminary work by:

a. Developing polymerase chain reaction and pyrosequencing assays to quantify DNA methylation at selected CpG sites.
b. The application of bisulphite pyrosequencing assays to DNA samples of the discovery cohort.

Phase 2: To assess global methylation differences between tumour and adjacent tissue samples by:

a. Comparison of published modifications to the luminometric methylation assay to determine the optimum methodological approach.
b. Quantification of global DNA methylation levels in tumour and adjacent tissue samples of the discovery cohort.

Phase 3: To determine the presence of DMPs in a replication cohort of mixed colon and rectal tissue samples by:

a. Quantification of DNA methylation at selected CpG sites in a larger sample cohort of historically acquired colon and rectal tissue samples (n=68).

b. Determine the relationship between DNA methylation changes observed between matched tissue samples, and tumour cell content of the sample analysed.

Phase 4: To assess the relationship between DNA methylation changes and clinicopathological features by:

a. The retrospective collection of patient demographics and tumour features of samples obtained.
b. The comparison of DNA methylation profiles observed between samples of differing clinicopathological subgoups.
c. To use data of multiple DMPs to develop models for the prediction of tumour characteristics on the basis of DNA methylation.

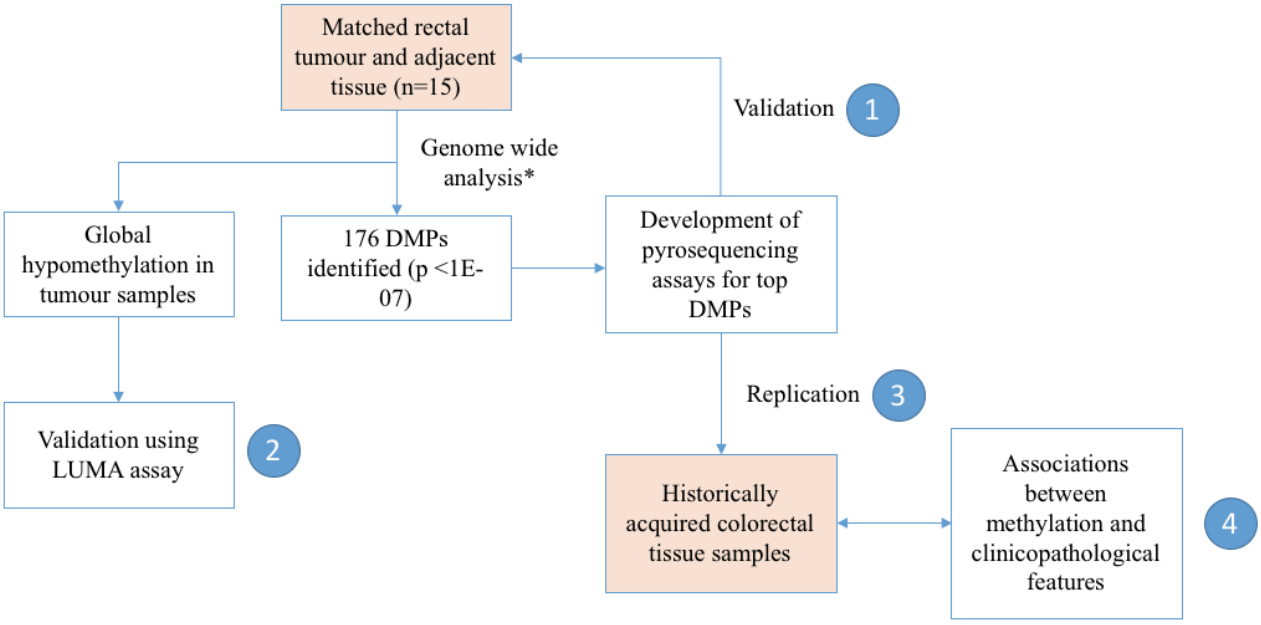


Figure 3.1: Overview of study design. Aims of the study are indicated by blue circles. * depicts preliminary work. DMP, differentially methylated probe; LUMA, luminometric methylation assay. Shaded boxes indicate tissue sample cohorts used for analysis.

4 Methods 1: Materials and Protocols

4.1 Administration

4.1.1 Exeter NIHR Clinical Research Facility Tissue Bank

This work was conducted in partnership with the Exeter NIHR Clinical Research Facility Tissue Bank (ETB). The ETB works within the National Institute for Health Research (NIHR) funded Exeter Clinical Research Facility (CRF). Its role is to collect and store tissue obtained from consenting patients for research purposes. Ongoing support was provided by the ETB team throughout the study.

4.1.2 Ethical approval

The ETB and CRF possess ethical approval for the collection and storage of human tissue for use in genetic and epigenetic research projects. Projects were approved by The Royal Devon and Exeter Tissue Bank Steering Committee.

4.1.3 Data collection and storage

All biological samples were stored in line with guidelines described by the Human Tissue Act and Medical Research Council (MRC).

Tissue samples were distinguishable only by unique tissue bank and study numbers. Hard copies of signed consent and data collection forms were stored in a study-specific folder within the ETB. Extracted data was also stored on password protected computers within the ETB. Patient identifiable data was accessible only to members of the ETB team, and to members of the study team on request. Archiving was undertaken as per standard Royal Devon & Exeter NHS Foundation Trust (RD&E) protocols.

4.2 Tissue samples

4.2.1 Patients

This study used tissue samples obtained from two distinct collection periods as follows:

RIST samples:

RIST samples were collected as part of an ongoing, prospective cohort between March 2015-April 2016. Full details of the RIST study are described in Section 11.7.

Patients older than 18 years with known rectal cancer treated surgically with curative intent at the RD&E were recruited to this study. Rectal cancer was defined as histologically proven carcinoma in which the lower edge of the tumour was situated less than 15cm from the anal verge on mid-sagittal MRI. Patients under the age of 18 years, pregnant and/or unable to provide informed consent were excluded from recruitment. Eligible patients were identified by treating clinicians and invited to participate in the study by a member of the ETB (see “Tissue acquisition” below, Section 4.2.2).

Historically acquired samples:

Collected between 2004-2007 and stored by the ETB.

Patients older than 18 years with known colorectal cancer treated surgically with curative intent at the RD&E were recruited to this study. Histologically proven carcinoma situated anywhere within the colon or rectum were included in this cohort. Patients under the age of 18 years, pregnant, and/or unable to provide informed consent were excluded from recruitment.

4.2.2 Tissue acquisition

Informed consent was obtained from each patient by a qualitied member of the ETB team prior to surgery in both cohorts.

RIST samples

Rectal tissue biopsies were taken at the time of defunctioning stoma formation, endoscopy or definitive surgery by the surgeon performing the procedure. Two biopsies were taken from each tumour and adjacent mucosa. Non-malignant tissue was identified visually at a distance determined by the operating surgeon. In cases where biopsy was not possible or contraindicated, samples were dissected by a consultant histopathologist following removal of the entire surgical specimen. Blood samples (plasma, serum and PAXgene blood RNA) were also obtained by the anaesthetist at the time of surgery and will be used in future studies to investigate epigenetic biomarkers. Samples obtained from each patient were allocated and labelled with unique tissue bank and study numbers.

Tissue samples were transferred immediately to the CRF situated on the RD&E site, where they were flash frozen in liquid nitrogen and stored at -80℃ for future use.

Historical samples

Flash-frozen, matched colorectal tissue samples were identified from a database of historically acquired specimens stored by the ETB. Samples were selected to provide an even distribution of left and right sided tumours.

An application was made to the ETB for the use of these samples (Approval no. STB41 / CRF266). Samples were transferred to the epigenetics research laboratory freezers situated in the same room as the ETB freezers with care taken to minimise transfer time and therefore avoid thawing.

FFPE samples

An additional subset of historically acquired samples had been selected in preliminary work and preserved by formalin fixation and paraffin embedding (FFPE). DNA was extracted in preliminary work using the QIAamp DNA FFPE Tissue Kit (Qiagen) kit, however this failed to meet the required quality for genome wide analysis. These samples will be referred to as the FFPE cohort for the remainer of this thesis and will be discussed only in Section 10. All other reference to historically acquired samples will refer specifically to those preserved by flash-freezing.

4.2.3 Clinical and demographic data

Demographic information was collected at the time of consent (age, gender and ethnicity). Additional clinical information was obtained retrospectively from the RD&E's CDM and PACs imaging software.

Clinic letters, multidisciplinary team meeting summaries and operation notes were reviewed using the CDM system. The Path system was used to review histopathology reports. With regard to historically acquired samples, five-year outcome data was available. However due to the retrospective nature of the study, and lack of accessibility to death certificates, accurate data regarding cause of death, time to relapse and disease-free survival were not obtainable. Five-year overall survival (OS) was therefore the only outcome considered appropriate for inclusion. Five-year OS was defined as death from any cause within or equal to 60 months of specimen retrieval.

Clinicopathological features of RIST samples

Samples analysed in our preliminary studies were obtained from the RIST cohort. The clinicopathological features of these samples are summarised in Table 4.1.

Rectal adenocarcinoma was diagnosed in 13/15 patients. The remaining two patients were diagnosed with tubulovillous adenoma

with high grade dysplasia (R05 and R06). Three tissue sample pairs were obtained from patients prior to neoadjuvant chemoradiotherapy (nCRT, Section 11.7). In these cases, biopsies were taken at the time of endoscopy where no assessment of EMVI, Dukes' stage, T stage or N stage were reported. Two of these patients (R10 and R13) achieved complete pathological response following nCRT and so this data remains unavailable.

Characteristic	Total n=15
Mean age (± SD)	68.7 ± 10.8
Sex	
Male	6
Female	9
Tumour type	
Adenocarcinoma	13
Tubulovillous adenoma with high grade dysplasia	2
Tumour site	
Low rectum	8
Mid rectum	2
High rectum	5
Differentiation	
Well/Moderate	11
Poor	1
NA/NR	3
Dukes' stage	
A	3
B	2
C	5
NA/NR	5
Pathological T stage	
1	2
2	2
3	6
NA/NR	5
Pathological N stage	
0	5
1	5
NA/NR	5
EMVI	
Positive	3
Negative	7
NA/NR	5

Table 4.1: Summary of clinicopathological features of the RIST cohort. SD, standard deviation; NA, not applicable; NR, not reported.

Clinicopathological features of historically acquired samples

Table 4.2 summarises patient demographics and tumour features of the historically acquired sample cohort.

Characteristic	Total n= 68
Sex	
Male	35
Female	33
Mean age (± SD)	72.7 (± 10.0)
Tumour site	
Left	35
Rectum	9
Rectosigmoid/ sigmoid	21
Descending colon	5
Right	33
Ascending and transverse colon	14
Caecum	19
Histological tumour type	
Adenocarcinoma	65
Mucinous Adenocarcinoma	3
Duke's stage	
A	5
B	45
C	16
D	2
Pathological tumour (T) stage	
1	1
2	7
3	49
4	11
Pathological node (N) stage	
0	49
1	16
2	6
Metastasis	
Yes	2
No	66
Extramural vascular invasion	
Yes	10
No	57
NR	1
Differentiation	
Well/Moderate	64
Poor	4

Table 4.2. Summary of patient and tumour characteristics of the historically acquired cohort. SD, standard deviation; NR, not reported.

4.3 Tissue dissection and DNA/RNA extraction

4.3.1 Introduction

Genomic DNA and total RNA were each extracted from historical tissue samples using the Qiagen AllPrep DNA/RNA Mini Kit. In this protocol tissue disruption and homogenisation are performed simultaneously to ensure the respective release and efficient binding of nucleic acids. DNA from RIST tissue samples had been extracted in preliminary studies using the same kit.

A guanidine-isothiocyanate–containing buffer inactivates DNases and RNases during homogenisation and therefore prevents digestion of DNA and RNA respectively. As the lysate passes through an AllPrep DNA spin column, the high-salt buffer promotes the selective binding of DNA which is then washed and eluted.

RNA is isolated from the flow through of the AllPrep DNA column. Ethanol promotes the binding of RNA to the RNeasy spin column which is again washed and eluted with RNase free water.

4.3.2 Tissue sample preparation and histological assessment

Historical tissue samples were transferred from the -80℃ freezer to the laboratory on dry ice to avoid thawing. Up to 30mg of frozen tissue was dissected macroscopically for DNA/RNA extraction, using a fresh scalpel and Petri dish on dry ice. The mass of each sample was weighed prior to homogenisation to ensure that 30mg limit was not exceeded.

Immediately adjacent tissue to that extracted from, was dissected and fixed in formalin prior to paraffin embedding (Figure 4.1). Histological assessment was performed by a consultant histopathologist and reported in increments of 5%.

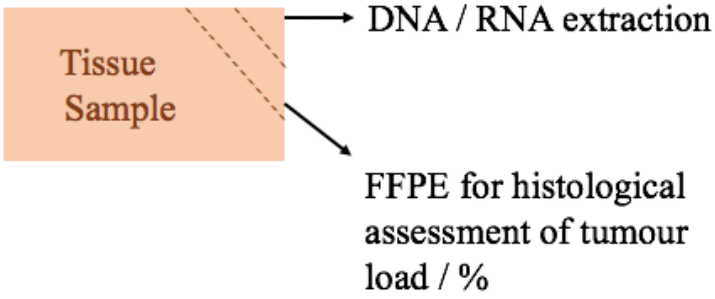


Figure 4.1: Schematic diagram to illustrate the way in which adjacent tissues were prepared for histological assessment and DNA/RNA extraction. FFPE, formalin fixed paraffin embedded.

4.3.3 DNA/RNA extraction protocol

The manufacturer's protocol was followed as outlined in Figure 4.2.

10 µl of β-ME was added per 1 ml Buffer RLT Plus prior to homogenisation. Work surfaces and pipettes were cleaned with RNaseZap (Applied Biosystems) prior to use and filter pipette tips were used throughout.

Disruption and homogenisation was completed using a glass homogeniser and 600ul Buffer RLT Plus/ β-ME solution (Sigma Life Sciences). Homogenised lysate was transferred into 1.5ml ependorph tube. Additional disruption was completed using a rotary homogeniser and disposable tips where necessary.

Solid material was separated from the remaining lysate by centrifuge at 10,000 x g for three minutes. The supernatant was transferred to an AllPrep DNA spin column and centrifuged for 30 seconds. The AllPrep DNA spin column was placed into a new 2ml collection tube and stored at room temperature throughout the RNA extraction period. The resultant flow though was used for RNA extraction.

600ul of 70% ETOH was added to the RNA-containing flow through and mixed by pipetting. 700ul of the solution was added to a RNeasy spin column and centrifuged for 30 seconds. The flow through was disposed of and process repeated using the remaining RNA-containing solution. 700ul and

500ul of buffers RW1 and RPE respectively were passed though the column at 10000 x g for 15 seconds each. An additional two-minute wash was performed using 500ul of RPE, followed by a one minute “dry” spin in a clean collection tube to minimise residual ethanol contamination. RNA was eluted using 40ul RNase free water into a sterile collection tube. This step was repeated to produce a final volume of 80ul. Eluted RNA was transferred into labelled Matrix collection tubes and immediately transported to a -20℃ freezer for storage.[3]

The AllPrep DNA spin column was washed in two steps using 500ul of AW1 and AW2 for 15 seconds and two minutes respectively. DNA elution was completed in two steps of 75ul EB buffer each (150ul total). DNA was transferred into labelled Matrix collection tubes and stored at 5℃ for 24-48 hours prior to spectrophotometry and dilution.

[3] RNA extracted from these samples was not analysed in the current study. Instead, it was transferred to a -80˚C freezer for storage and will be utilised in future work. The extraction process has been included for the sake of completion.

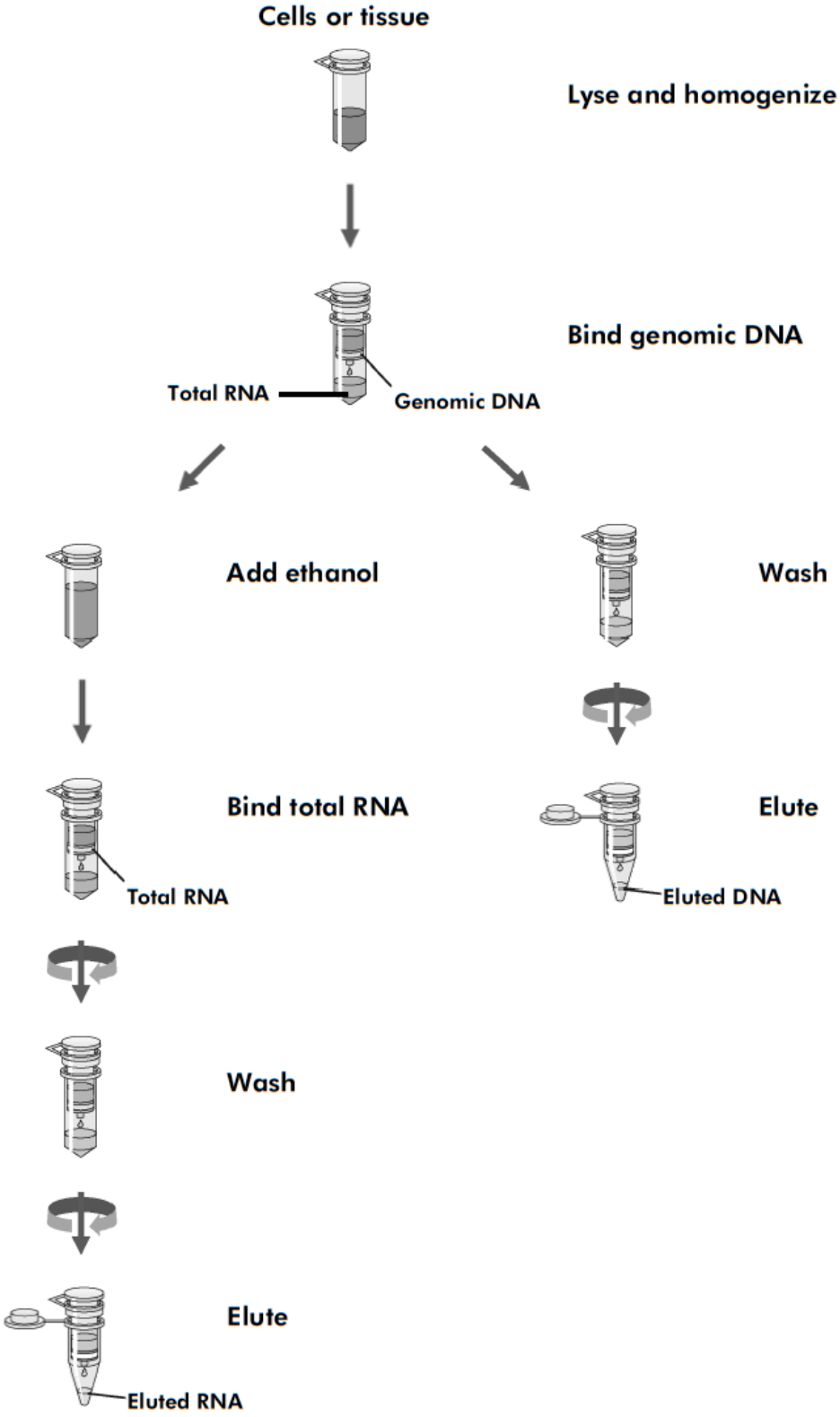


Figure 4.2: Overview of Qiagen AllPrep DNA/RNA Mini Kit protocol.

4.4 Spectrophotometry

4.4.1 Introduction

The Thermo Scientific NanoDrop spectrophotometer was used to determine the quality and quantity of DNA obtained from extraction (Figure 4.3). This method provides a simple, reproducible and accurate estimation of DNA concentration and purity.

DNA concentration is measured based on the absorbance at 260nm and the selected analysis constant.(144) A modified Beer-Lambert equation is used to correlate the calculated absorbance with concentration.

$$C= (A*e)/b$$

Where:

c= Nucleic acid concentration in ng/l

A= Absorbance in AU

e= Wavelength-dependent extinction coefficient in ng-cm/µl (50 for ds-DNA)

b= Path length in cm

Assessment of DNA quality is required in order to determine the suitability for further assessment. The purity of DNA is quantified using the ratio of absorbance at 260 and 280 nm. A 260/280 ratio between 1.7-2.0 is generally considered pure. Protein, phenol from DNA extraction and other contaminants each absorb strongly at/around 280nm. The presence of such contaminants therefore, results in a low 260/280 ratio. An additional absorbance ratio of 260/230 provides a secondary measure of purity. 260/230 ratios below 1.8-2.2 indicate the presence of co-purified contaminants.

4.4.2 Protocol

The NanoDrop 8000 software (Thermoscientific) was used to interpret data. The nucleic acid measurement function was selected. Pedestals were cleaned using lint free tissues before use and between subsequent additions of reagents or DNA samples. The Nanodrop was zeroed with 1.5µl of deionised water as prompted. Blanking was performed using 1.5µl of the EB

elution buffer. Sample IDs were entered into respective wells and 1.5µl of DNA solution was added to each pedestal for measurement.

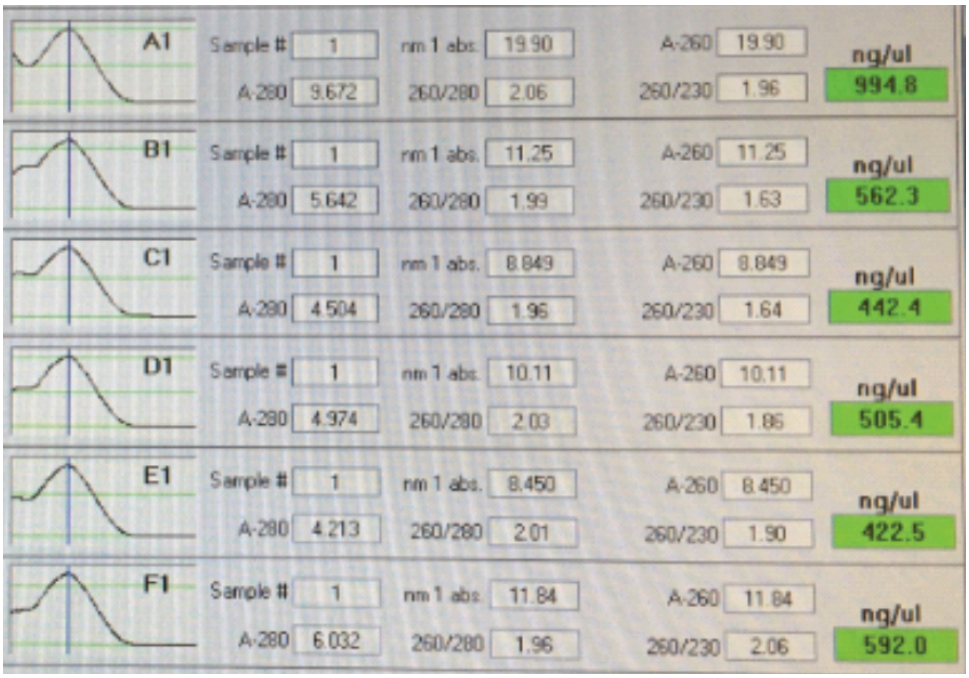


Figure 4.3: Nucleic acid quality and quantity as determined by NanoDrop 8000 software for six samples. The purity of DNA is quantified using the ratio of light absorbance at wavelengths of 260nm and 280nm. A 260/280 ratio between 1.7-2.0 is generally considered pure. Green boxes indicate the concentration of DNA analysed.

4.5 DNA dilution

DNA elutions of concentration greater than 250ng/µl were manually diluted to approximately 200µl in EB buffer prior to robotic dilution. The MultiProbe II Plux HT EX Robotic Liquid Handling System (Perkin Elmer) Robot was then used to accurately pipette 20µl of DNA and dH_2O into each well of a 96-well plate at a concentration of 25ng/µl.

4.6 Bisulphite Conversion

4.6.1 Background

Sodium bisulphite conversion involves the selective conversion of unmethylated cytosine nucleotides to uracil. Methylated cytosines remain unconverted. The process therefore enables the distinction between individual cytosines and methylcytosines to be made via downstream sequencing techniques (Table 4.3).

	Original sequence	After bisulphite treatment	After PCR
Unmethylated DNA	A-C-G-T-C-G-T-C-A	A-U-G-T-U-G-T-U-A	A-T-G-T-T-G-T-T-A
Methylated DNA	A-C-G-T-C-G-T-C-A	A-C-G-T-C-G-T-U-A	A-C-G-T-C-G-T-T-A

Table 4.3. The effect of bisulphite treatment and amplification on DNA sequence. Changes are highlighted in red. PCR, polymerase chain reaction. Adopted from (145)

Bisulphite conversion was performed using the EZ-96 DNA Methylation-Gold™ Kit (Zymo Research, D5007). This kit uses temperature denaturation to permit the simultaneous denaturation and bisulphite conversion. The process allows efficient conversion of >99% of non-methylated cytosines to uracil and <1% of methylated cytosines.(146) DNA recovery is reported to be greater than 75%.

4.6.2 Protocol

The conversion process was carried out as per the manufacturer's instructions with minor modifications as outlined (Figure 4.4).(146)

9ml of dH_2O, 500µl of M-Dissolving Buffer and 3ml of M-dilution Buffer were added to the CT Conversion Reagent bottle provided and mixed with frequent vortexing for 15 minutes. 144ml of 100% ethanol was added to the 36ml M-wash buffer prior to use.

130µl of CT conversion reagent was added to 20ul of DNA (25g/µl) in a 96-well plate and mixed by pipetting. The plate was sealed with an adhesive PCR plate seal (ThermoFisher Scientific) and transferred to a thermal cycler set to 98℃ for ten minutes, 64℃ for 2.5 hours and 4℃ for up to 20 hours.

400µl of M-Binding Buffer was added to a Silicon-A™ Binding Plate mounted on a Collection Plate provided. 150µl of DNA/Conversion reagent solutions were added and mixed by pipetting. The plate was then centrifuged for 5 minutes at 3000 x g and the flow-through was discarded.

400µl of M-Wash Buffer was added to each well and centrifuged for a further five minutes. 200µl of M-Desulphonation Buffer was added to each well and incubated at room temperature for 25 minutes. Following incubation, the plate was centrifuged for five minutes. Two additional wash cycles were performed with 400µl of M-Wash buffer and centrifugation for five minutes and ten minutes.

25µl of M-Elution buffer was added to the Silicon-A™ Binding Plate on an elution plate provided. Following incubation at room temperature for 5 minutes, the plates were centrifuged for three minutes. This step was repeated to produce a final volume of 50µl per well. Converted DNA was transferred to matrix tubes and stored at -20℃ for storage.

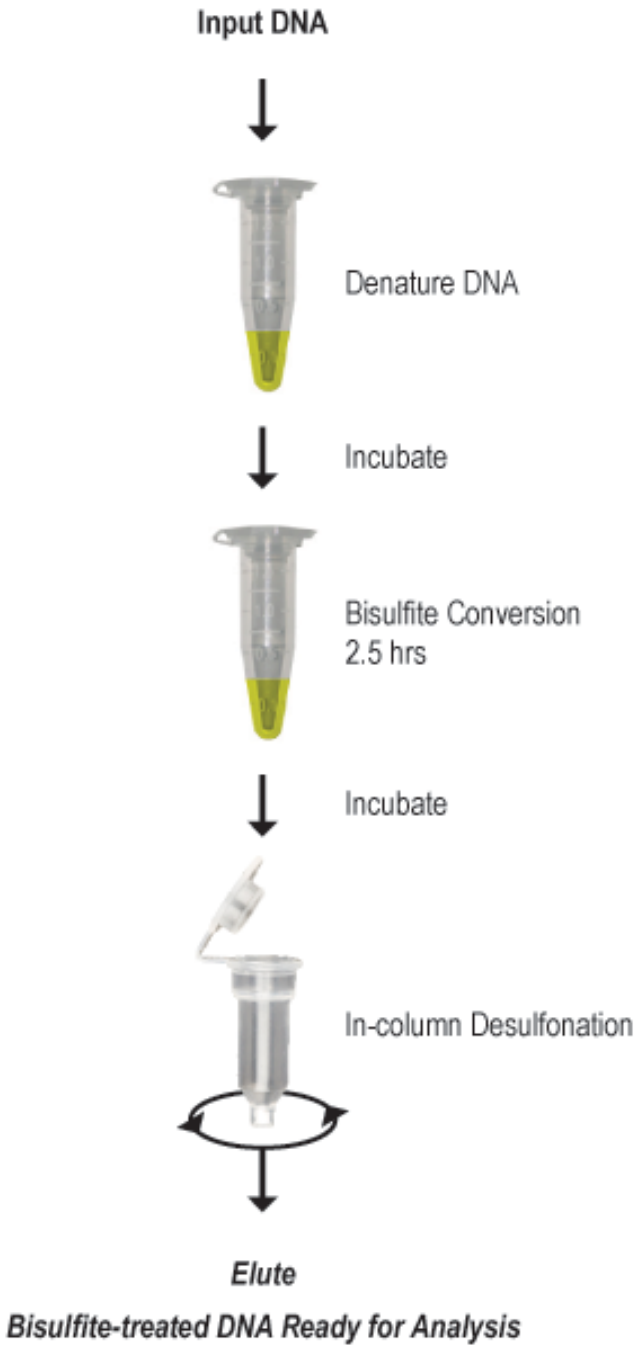


Figure 4.4: EZ-96 DNA Methylation-GoldTM Kit (Zymo Research, D5007).

4.7 Polymerase Chain Reaction

4.7.1 Background

Polymerase chain reaction (PCR) was required to produce sufficient DNA for pyrosequencing analysis. PCR allows the exponential amplification of genomic DNA.(147) The process is catalysed by the enzyme *Taq* polymerase, which remains stable at the high temperatures required for DNA denaturation. Primers are short sequences of nucleotides which act as a starting point for DNA synthesis by *Taq* polymerase. Specific primers are required to complement the 3' end of each strand.

Figure 4.5 illustrates the steps involved in PCR. Of note, PCR of bisulphite converted DNA results in the substitution of uracil nucleotides with thiamine. Final volumes of PCR reagents are shown in Table 5.5.

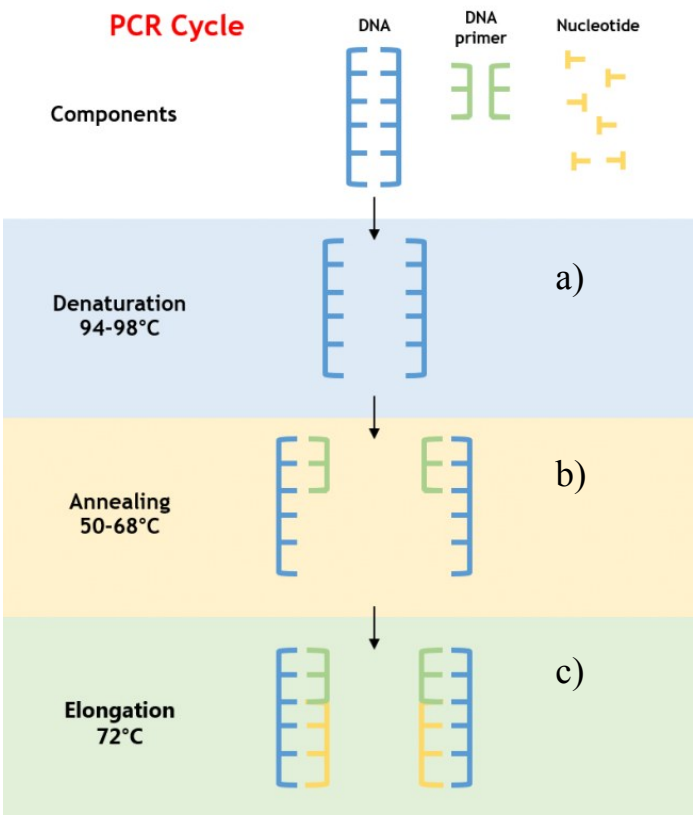


Figure 4.5: Schematic diagram to illustrate the steps of PCR.
a) Heating of DNA results in the disruption of hydrogen bonds and subsequent separation of DNA strands.
b) Cooling allows the annealing of primers to the single stranded DNA templates. The annealing temperature is specific to the individual primers used.
c) Complimentary nucleotides are added to the template from 5' to 3' through the condensation of nucleotide 5'-phosphate groups with 3'hydroxyl groups of the extending strand. Figure from Ref (147)

4.7.2 Primer design

500 base pairs upstream and downstream to the CpG of interest were identified using UCSC genome browser (Feb. 2009 GRCh37/hg19. human assembly. http://genome-euro.ucsc.edu/cgi-bin/hgGateway?clade=mammal&org=Human&db=hg19&redirect=manual&source=genome.ucsc.edu) and copied into PyroMark Assay Design 2.0 software.(148) The bisulphite conversion function was used to replace non-CpG cytosine nucleotides with thymine and to highlight potentially methylated

CpGs. Primers for amplification of each DMP were designed with the following principles in mind.

Potentially methylated CpG dinucleotides and repetitive sequences were avoided to prevent selective annealing of methylated/unmethylated CpGs and mispriming respectively (Figure 4.6). Where possible, the melting temperatures of each forward and reverse primer pair were kept close and within the 55-65℃ range. Predicted primer performance was used as a guide only and a low score did not exclude a primer set from further inclusion.

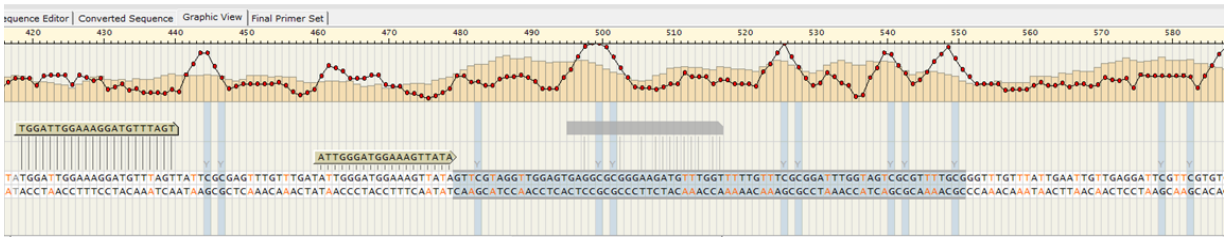


Figure 4.6: Screenshot demonstrating forward and sequencing primer design using the PyroMark Assay Design 2.0 software. The reverse primer is omitted from the figure. The pyrosequencing target sequence is highlighted in blue. The CpG of interest is located at position 501.

The accuracy of pyrosequencing is diminished by strings of repetitive single nucleotides and with increasing distance from the sequencing primer. Sequencing primers were therefore designed in close proximity to the CpG of interest and single nucleotide repeats were avoided where possible. The primer on the complementary strand to the sequencing primer was biotinylated to facilitate immobilisation to streptavidin beads during pyrosequencing.

Primers were ordered from Integrated Genome Technologies and resuspended to 100μM in sterile water. Working dilutions were made by diluting 20μl of forward and reverse primer in 200μl of total volume (10μM forward reverse primer mix). Sequencing primers were diluted separately to 10μM. Primers were stored at -20℃ for future use.

4.7.3 PCR Protocol

PCR was performed in an AirClean® Systems AC600 Series PCR Workstation (ThermoFisher) following 30 minutes of ultra-violet irradiation. Empty PCR plates and Ependorph tubes were held in the workstation throughout the irradiation process. All DNA, primers and reagents (except *Taq* polymerase) were thawed completely at room temperature and vortexed prior to use. Taq polymerase was stored on ice throughout the process. Sterile filtered pipette tips were used to minimise contamination.

1-4µl of DNA was initially transferred to each well of a 96-well PCR plate depending upon the efficiency of the PCR and pyrosequencing primers used (see Table 5.5, Section 5.3 for optimised PCR reaction conditions). All other reagents required for PCR were combined to create master mixes of sufficient volume in 1.5ml ependorph tubes. Master mixes were mixed by vortexing prior to addition of *Taq* polymerase. Appropriate volumes of master mix (final PCR volume – DNA volume) were added to the DNA of each well and mixed by gentle pipetting. Bisulphite converted 100% methylated HeLA DNA (New England BioLabs) was used as a control. Each PCR plate was sealed with an adhesive PCR plate seal prior to thermal cycling.

The Applied Biosystems Veriti^{TM} 96-Well Thermal Cycler was used to perform thermal cycling. Temperatures and times of each stage are displayed in Table 4.4. 40 cycles were performed in all instances with the aim to maximise PCR product without excessive mutation rate.

PCR optimisation and final reaction conditions are discussed fully in Section 5.

Step	Time/mins	Temperature/ ˚C
Initiation	15.0	95.0
Denaturation	0.5	95.0
Annealing	0.5	Variable
Extension	1.0	72.0
Final extension	10.0	72.0
Holding	∞	15.0

Table 4.4: Summary of time and temperature conditions used for PCR. Specific annealing temperatures are listed in Table 5.5.

4.8 Agarose gel electrophoresis

4.8.1 Background

Agarose is a linear polymer of galactose which when dissolved in Tris/Borate/EDTA (TBE), forms a gel.(149) Tris base and boric acid solution acts as a pH buffer, whereas EDTA protects nucleic acids from enzymatic degradation through the chelation of magnesium ions. When an electric current is passed through the agarose gel, nucleic acid fragments are separated by molecular size. The presence of negatively charged phosphate groups results in the migration of DNA fragments from the negative to positive electrodes. Syto60 is a fluorescent nucleotide stain which enables the visualisation of DNA fragments following electrophoresis. A ladder, which comprises DNA fragments of specific sizes, is used to estimate the size of PCR products (Figure 4.7).

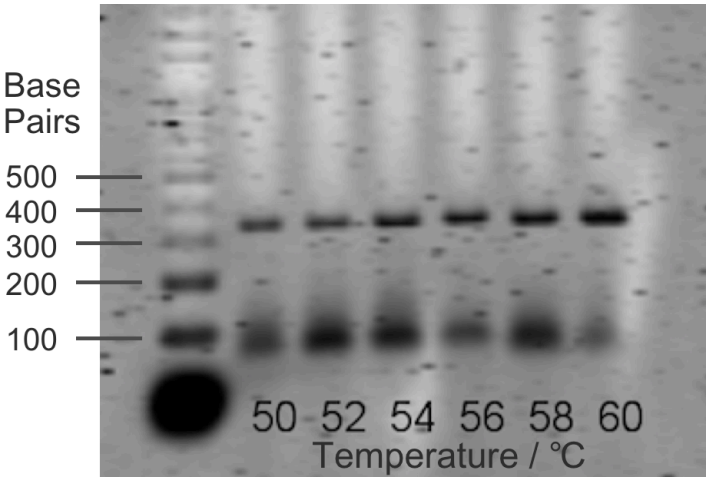


Figure 4.7: Example of agarose gel electrophoresis using a 338bp PCR product at temperatures 50-60°C. The 100bp ladder on the left enables estimation of amplicon size.

Two protocols were used in this study. Amplified bisulphite converted DNA were assessed using protocol 1 following PCR. Protocol 2 was used to assess the integrity of untreated genomic DNA prior to global DNA methylation analysis.

4.8.2 Protocol 1 (Syto60-stained)

1.5g agarose powder (Sigma) was added to 100ml TBE (89mM Tris (pH 7.6); 89mM boric acid; 2mM EDTA) in conical flask and heated using a microwave until all agarose powder was dissolved. 10µL Syto60 was added to the

agarose-TBE solution and allowed to set with 24 slot comb in situ to allow the formation of loading wells.

The gel was submerged in TBE solution within an electrophoresis tank.

5µL of PCR product was added to 2µL a loading buffer (Orange G, 0.05g; FICOLL 400, 1.5g; 0.5M EDTA pH8, 1ml; dH_2O, 9ml) and mixed by pipetting in a clean PCR plate. 5µl of PCR product-loading buffer solution was added to each well formed in the agarose gel. A 100-base-pair ladder (New England BioLabs) was added to the first well of each row to enable quantification of amplicon size. The presence, size and intensity of each PCR product was visualised using the LI-COR Image Studio Lite software and LI-COR Odyssey® CLx scanner. Darker bands correspond to higher volumes of PCR product.

4.8.3 Protocol 2 (ethidium bromide-stained)

1.3g agarose powder (Sigma) was added to 150ml TBE (89mM Tris (pH 7.6); 89mM boric acid; 2mM EDTA) in conical flask and heated using a microwave until all agarose powder was dissolved. 0.9µL of ethidium bromide solution was added to the agarose-TBE solution and allowed to set with 24 slot comb in situ to allow the formation of loading wells.

The gel was submerged in TBE solution within an electrophoresis tank.

100ng of genomic DNA was added to 2µL a loading buffer (Orange G, 0.05g; FICOLL 400, 1.5g; 0.5M EDTA pH8, 1ml; dH_2O, 9ml) and mixed by pipetting in a clean PCR plate. 5µl of DNA-loading buffer solution was added to each well formed in the agarose gel. A 1 kilobase-pair ladder (New England BioLabs) was added to the first well of the row to enable quantification of amplicon size. DNA was visualised and photographed under ultraviolet light.

4.9 Pyrosequencing

4.9.1 Background

Bisulphite pyrosequencing was used to quantify DNA methylation at individual CpG dinucleotides. Pyrosequencing is a method of sequencing by synthesis. The principles of its use are illustrated in Figure 4.8. DNA methylation level is determined by the relative proportion of cytosine to thiamine peaks at potentially methylated CpGs.(150)

4.9.2 Assay design

Pyrosequencing assays were designed using the PyroMark Assay Design software version 2.0 (Qiagen). The target sequence previously identified during the primer design stage, was pasted into sequence generator. The dispensation sequence suggested by the PyroMark software was modified to include bisulphite controls and additional nucleotides in repetitive regions (Figure 4.9).

4.9.3 Preparation

Enzyme and substrate were each suspended in 620µL of fresh dH20. Three columns were cut from a 96-well PCR plate. 50ml of wash buffer, denaturing solution, dH20 and wash buffer were added to each trough of the wash station as indicated. 50ml of dH20 was vacuumed prior to use on each occasion to avoid contamination.

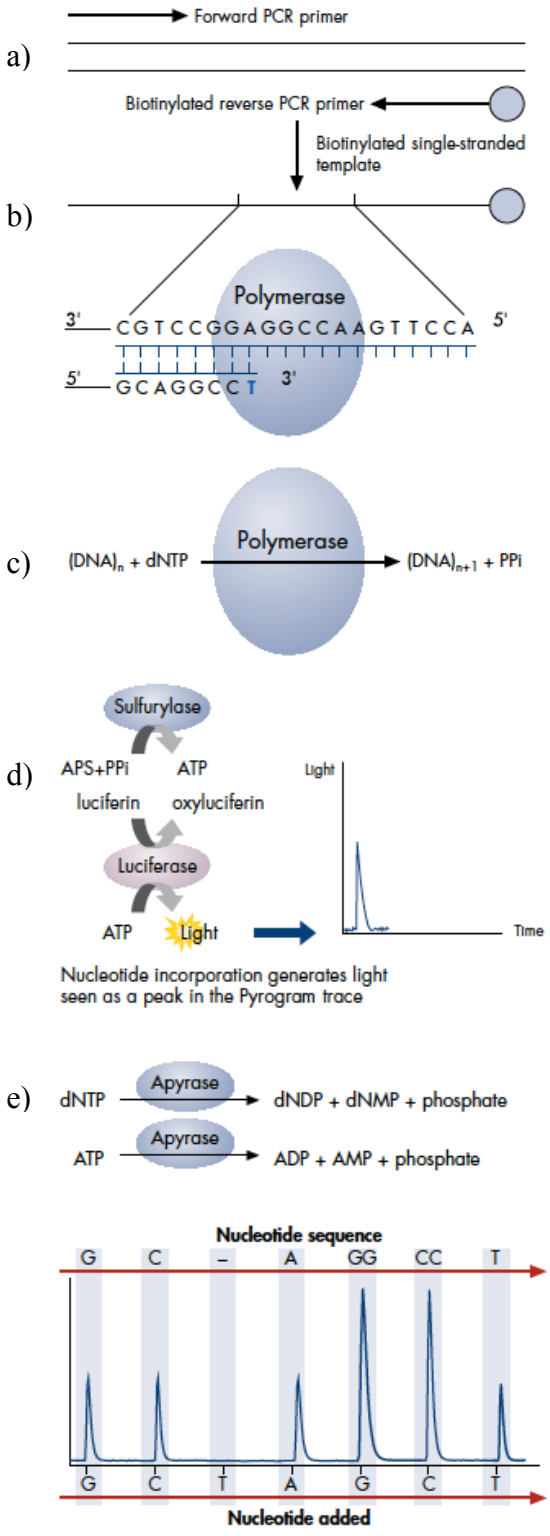


Figure 4.8: Principles of pyrosequencing.

a) The biotinylated strand of PCR product is isolated by denaturation and hybridized with a sequencing primer.

b) Primer-template hybrids are incubated with DNA polymerase, ATP sulfurylase, luciferase, and apyrase, as well as the substrates adenosine 5' phosphosulfate (APS) and luciferin

c) The incorporation of complimentary dNTPs results in the proportional release of pyrophosphate (PPi) molecules.

d) PPi is converted to ATP in the presence of APS and ATP sulfurylase. ATP provides the energy to convert luciferin to oxyluciferin. This reaction produces a proportional amount of light which is detected by the pyrosequencer and represented as peaks on a pyrogram.

e) Apyrase degrades unincorporated nucleotides and ATP between subsequent additions of dNTP.

Taken from Ref (145)

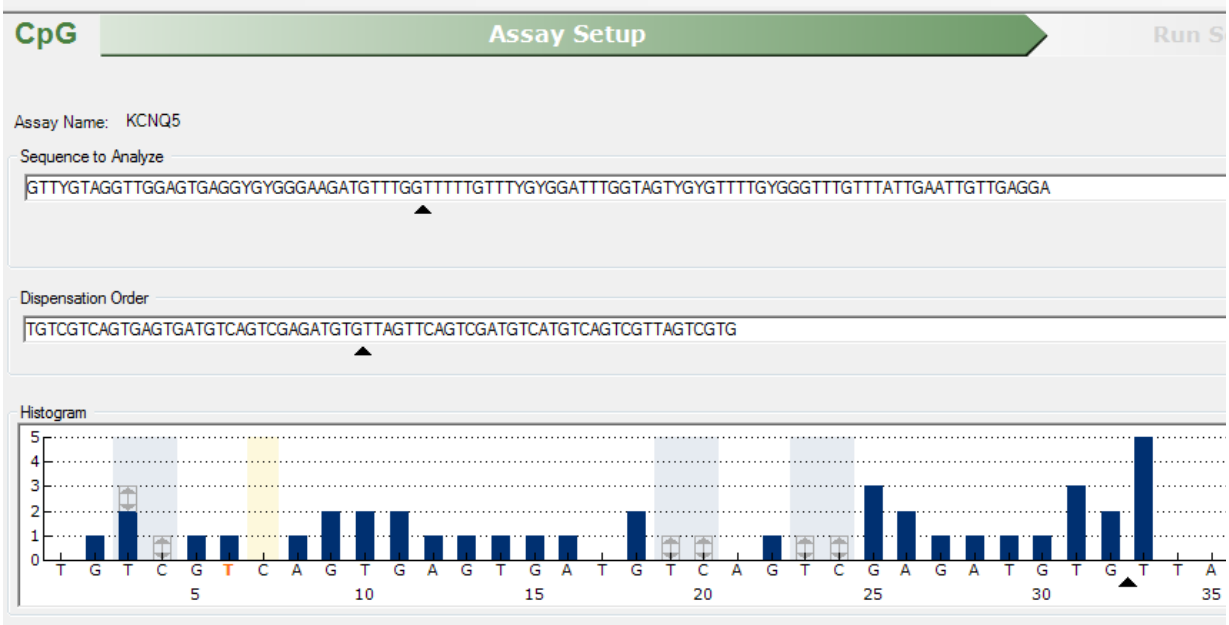


Figure 4.9: Example of a pyrosequencing assay designed using the PyroMark Assay Design software version 2.0 (Qiagen). The top sequence represents the sequence to be analysed and includes the target CpG. The bottom sequence represents the dispensation order of nucleotides. At the position indicated by the arrows, two thiamine nucleotides are dispensed to overcome the potential refractory period of polymerase caused by multiple thiamine repeats. A bisulphite control and three CpG dinucleotides are highlighted in yellow and blue respectively.

4.9.4 Pyrosequencing protocol

Streptavidin beads (Qiagen) were gently shaken until a homogenous solution was formed. 2µl of streptavidin beads and 40µl of binding buffer (Qiagen) per sample were used. In each instance a sufficient volume of bead/buffer mix was made for 28 reactions (56µl beads, 1120µl binding buffer) in a 1.5ml ependorph tube. 42µl of bead/buffer mix was added to each well of the 3 column plate prepared earlier. Between 15-25µl of PCR product was added to each well depending upon the efficiency of the individual PCR/pyrosequencing primer set (see Tables 5.4 and 5.5 for optimum pyrosequencing conditions). dH20 was added to make the final volume equal

to 80µl. An adhesive PCR lid was used to seal the plate before placing onto a shaker for 10 minutes.

0.85µl of sequencing primer and 24.15µl of annealing buffer per sample were used. In each instance a sufficient volume of primer/buffer mix was made for 28 reactions (23.8µl sequencing primer and 676.2µl annealing buffer) in a 1.5ml ependorph tube. 25µl of primer/buffer mix was added to each well of a 24-well pyrosequencing plate and placed on the indicated position of the workstation.

The DNA/bead/buffer plate was moved immediately from the shaker to the indicated position on the workstation and lid removed. The vacuum filter-probes were placed into the wells until all sample was taken up. The probes were sequentially placed into the ethanol, denaturing and wash buffer troughs until liquid was seen to flow through the plastic tube for five, five and ten seconds respectively. The probes were removed from the wash buffer and held at 90' for an additional two seconds.

The vacuum was turned off and probes placed into the wells of the pyrosequencing plate. The beads were removed by gentle rocking within the primer/buffer solution. The plate was immediately transferred to an 80' heat block for 90 seconds and allowed to cool for five minutes before beginning the pyrosequencing run.

The cartridge was loaded as indicated by the Pyromark software.

4.9.5 Analysis of pyrosequencing data

All assays were performed with bisulphite converted DNA from corresponding tumour and adjacent mucosa tissue on the same PCR and pyrosequencing plates. Due to the large differences between tumour and adjacent tissue, duplicate analyses were not performed as a matter of routine.

Quality control (QC) was performed by the Pyromark software using pre-determined tolerances set by the manufacturer. The software compares theoretical peak heights calculated according to the sequence analysed, with those of non-CpG reference sites obtained from the sample analysis. The bisulphite control (Figure 4.7), which should theoretically achieve a peak height of zero, is also used in the QC process. CpG sites that passed the programme's internal QC were accepted. Low peak heights and/or the presence of background noise, result in uncertain or failed QC at one or more CpG sites. Where samples failed QC at the CpG site of interest, the assay was repeated using DNA of both tumour and adjacent mucosa. Where repeated assays failed the QC, for example due to very low peak height, both tumour and adjacent mucosa samples were excluded from analysis. It was felt that the high stringency of the internal QC was likely excessive in our study, as small inaccuracies in DNA methylation were dwarfed by large differences between tumour and adjacent tissue. As a result, CpG sites flagged as uncertain, and those that failed due to minor issues (e.g. failed bisulphite controls secondary to small amounts of background noise) were reviewed individually and repeated/excluded/included as appropriate.

4.10 Assessment of global DNA methylation: LUMA Assay

4.10.1 Background

The luminometric methylation assay (LUMA) was used to quantify global DNA methylation. This method was initially described by Karimi et al. and utilises the restriction endonucleases *Msp*I and *Hpa*II.(151) These isoschizomers each result in the digestion of DNA at 5'-CCGG-3' sequences to produce 5'-CG overhangs. Their action is determined however by the methylation status of the internal CpG dinucleotide.

The MspI enzyme is insensitive to CpG methylation and therefore results in cleavage of all CCGG sequences across the genome. In contrast, HpaII is inhibited by CpG methylation and therefore cleaves at unmethylated CCGG sites only (Figure 4.10). Quantification of 5'-CG overhangs is permitted by the addition of nucleotides during pyrosequencing. Each enzyme is used to digest DNA from the same sample in parallel reactions. The ratio of overhangs produced from each digestion is then used to calculate the percentage of global DNA methylation in each sample.

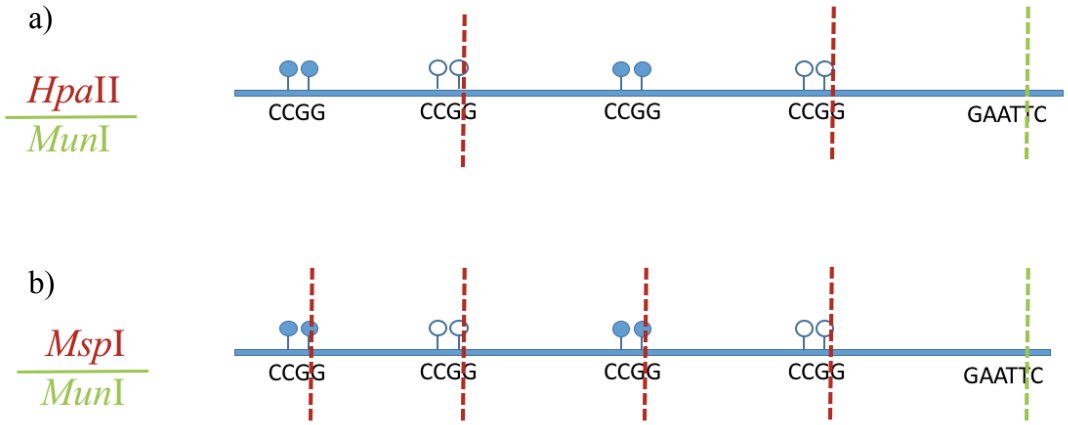


Figure 4.10: Schematic diagram to illustrate the methylation dependent action of a) HpaII and b) MspI restriction enzymes. The action of MunI (or EcoRI) is independent of methylation status and therefore acts to normalize DNA quantity.

The endonuclease MunI is also included in all restriction digests. MunI recognises the sequence GAATTC and is therefore uninfluenced by DNA methylation. As a result, the proportion of 5'-AATT overhangs produced remains consistent between reactions and thereby provides a reference for normalisation of DNA quantity. The restriction enzyme EcoRI exhibits similar action and is commonly used throughout the literature. The use of MunI was proposed by Lisanti et al. following the identification of star activity[4] of EcoRI.(153) The authors reported that although star activity could be reduced by increasing concentration of Tango buffer, the activity of HpaII/MspI was compromised by such endeavours.

LUMA assumes that all 5'-CG overhangs identified by pyrosequencing result from the enzymatic cleavage of HpaII and MspI. As a result, non-specific fragmentation of DNA may result in overestimation of HpaII activity and therefore the underestimation of global DNA methylation. Two methods have been described to address this concern. Duman et al. advocate the additional pyrosequencing analysis of undigested DNA, alongside digested samples, in order to quantify and correct for, the effects of fragmentation on a peak by peak basis.(154) However, this method poorly normalises DNA quantity between runs and demands considerably more DNA template than otherwise required. A more commonly adapted method was described by Bjornsson et al. (155) Their process involves the addition of GT repeats to the beginning of the pyrosequencing dispensation order in order to eliminate non-specific overhangs.

As a result, disparity of global methylation values is seen between this and other methods. Knothe et al. reported consistent discordance between methylation values obtained by the LUMA and LINE-1 assays which was emphasised amongst specific tissues. (156)

[4] Star activity is the phenomenon by which restriction enzymes exhibit relaxed specificity, i.e. identify sequences similar, but not identical, to their defined recognition sequences, under suboptimal conditions.(152)

4.10.2 Protocol

The LUMA assay was performed as described by Sant et al. with modifications as follows.(157)

Restriction digest

Master mixes were prepared for each digestion as follows:*

Mix A: 2 µL of 10× Tango Buffer™, 0.5 µL MunI (10 U/µL), and 0.5 µL HpaII (10 U/µL) per sample.

Mix B: 2 µL of 10× Tango Buffer™, 0.5 µL MunI (10 U/µL), and 0.5 µL MspI (10 U/µL) per sample.

3 µL of each mix was added to alternative wells of a 96-well PCR plate. The amount of sample required to yield 300 ng of DNA was added to each well and mixed by pipetting. DNase-free water was added to a final volume of 20 µL. Samples were immediately incubated at 37℃ for four hours in a thermal cycler.

Following incubation, 15µL of Annealing Buffer was added to each sample and mixed by pipetting. 17µL of Annealing Buffer/restriction digest solution was transferred to a Pyrosequencing plate in duplicate** (Figure 4.11).

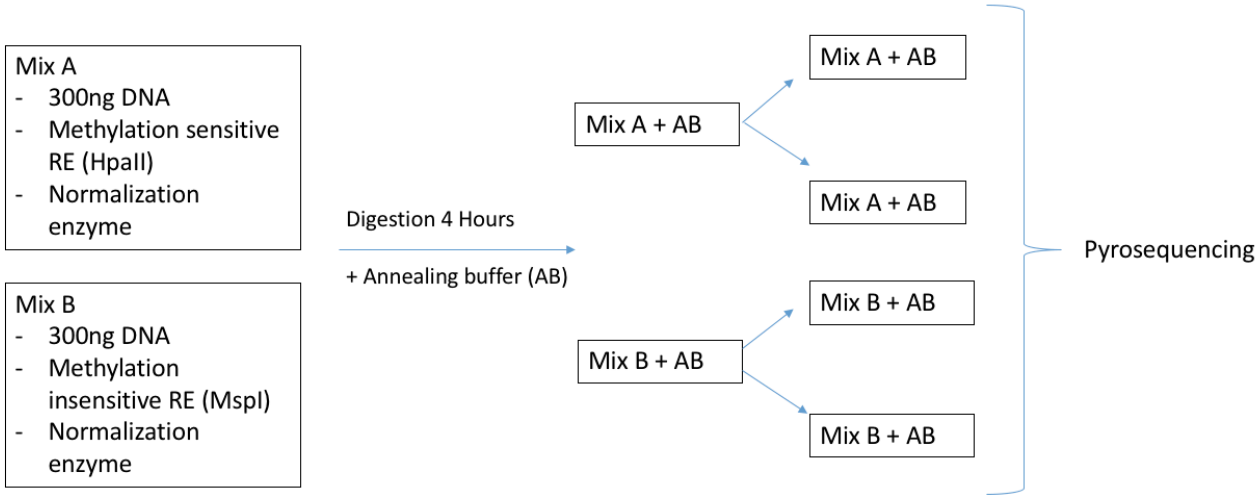


Figure 4.11: Flowchart of LUMA assay protocol. See text for full description.

Pyrosequencing of samples

Pyrosequencing assays were created using the PyroMark Q24 2.0.6 software (Qiagen) using the 'AQ Assay' (SNP) setup.

During assay optimisation, initial comparisons were made between the methodologies described by Karimi et al. and Bjornsson et al. (see Section 4.10).(151, 155) In order to achieve this, the order of nucleotides added by the pyrosequencer (dispensation order) to the primer-template hybrid solution were programmed as "ACTCGA" and "GTGTCACATGTGTG" respectively.(151, 155) These dispensation orders were used in separate assays to compare the global DNA methylation value recorded by each methodology.

The pyrosequencing cartridge was loaded as recommended by the PyroMark software prior to analysis.*** All duplicates and matched samples obtained from the same patient were analysed in parallel.

Analysis of data

Pyrograms produced from each assay were analysed using the PyroMark Q24 2.0.6 software (Qiagen). Mean peak heights were calculated from technical replicates. Global methylation level was determined using the following equations as described by Karimi et al. and Bjornsson et al. respectively:(151, 155)

For assays using the dispensation order "ACTCGA":

$$\left(1-\left(\frac{\text{HpaII (C+G)} / \text{MunI (Aa)}}{\text{MspI (C+G)} / \text{MunI (Ab)}}\right)\right) \text{x } 100$$

Where:

HpaII (C+G)= mean peak height of dispensation number 2 (Mix A)

MunI (Aa) = mean peak height of dispensation number 1 (Mix A)

MspI (C+G)= mean peak height of dispensation number 2 (Mix B)

MunI (Ab) = mean peak height of dispensation number 1 (Mix B)

For assays using the dispensation order “GTGTCACATGTGTG”:

$$\left(1-\left(\frac{\text{HpaII (G) / MunI (Ta)}}{\text{MspI (G) / MunI (Tb)}}\right)\right) \times 100$$

Where:

HpaII (G)= mean peak height of dispensation number 10 (Mix A)
MunI (Ta) = mean peak height of dispensation number 9 (Mix A)
MspI (G)= mean peak height of dispensation number 10 (Mix B)
MunI (Tb) = mean peak height of dispensation number 9 (Mix B)

* During LUMA optimisation, the performance of EcoRI and MunI was compared. In these reactions, 0.5 µL MunI (10U/ µL) was substituted for 0.5 µL EcoRI (10U/ µL) in each digestion mix. All other steps were unchanged.

** During LUMA optimisation, pyrograms and methylation values produced from the analysis of samples containing 300ng and 150ng of genomic DNA were compared. 300ng analyses were performed using all of the Annealing Buffer/restriction digest solution available (20 µL restriction digest + 15 µL Annealing Buffer). In these cases, analysis was performed in duplicate by preparing two initial solutions per restriction reaction. 150ng analyses were performed using 17 µL Annealing Buffer/restriction digest solution as described.

*** For dispensation order “ACTCGA”, nucleotides C and G were mixed and loaded together into position C of the pyrosequencing cartridge. Position G was loaded with dH_2O. Nucleotides were loaded as usual for dispensation order “GTGTCACATGTGTG”.

5 Methods 2: Assay Optimisation

5.1 PCR Optimisation

5.1.1 Primer design

In order to produce adequate quantities of DNA for bisulphite pyrosequencing, amplification of bisulphite converted DNA must first be performed by PCR. PCR requires high quality forward and reverse primers in order to replicate the specific amplicon of interest. Specific sequencing primers to the targeted sequencing region are also required for pyrosequencing.

The 176 DMPs identified via the Illumina Infinium 450k Beadchip assay were ranked by statistical significance. Primers were designed as outlined in Section 4.7.2. High quality primer design was not possible in all cases due to repetitive sequences and high CpG densities surrounding some DMPs of interest. As CpG sites undergo methylation sensitive conversion of cytosine to uracil during bisulphite treatment, primers overlapping CpGs may favour binding in one state or the other. In contrast, repetitive sequences interrupt the ability to produce primers specific to the site of interest. In these instances, the next most significant DMP was selected. Ten primer sets (forward, reverse and sequencing) were designed to assess eight hypermethylated and three hypomethylated DMPs (total n= 11). Coverage of both cg14650610 and cg24847829 within the *SPOCK_1* gene with a single primer set was possible due to their close proximity. DMPs selected for validation are shown in Table 5.1. All DMPs selected were within the top 24 most statistically significant.

In most cases, other CpGs in proximity to the DMP of interest were included within the 'target region' of the sequencing primer. For the remainder of this manuscript, DMPs selected for validation will be referred to by their associated gene name or chromosome number for clarity.

CpG ID	Array P Value	Array Δ / %	Chr	Position	Annotated gene	Gene group
cg14650610	7.30E-10	+48	5	136834492	*SPOCK1*	5'UTR
cg18538668	3.79E-09	-20	14	103839038		
cg24847829	6.29E-09	+40	5	136834464	*SPOCK1*	5'UTR
cg26034516	6.48E-09	+35	17	76228121	*LOC283999*	Body
cg03576469	6.83E-09	+29	19	46917061	*CCDC8*	TSS200
cg05447008	9.87E-09	+40	6	73331114	*KCNQ5*	
cg13356896	1.15E-08	+40	2	198650987	*BOLL*	
cg09129067	1.29E-08	-35	8	103750904		
cg24685755	1.39E-08	+35	19	53758031	*ZNF677*	5'UTR
cg10224098	2.26E-08	+40	1	44873229	*RNF220*	5'UTR
cg26314722	2.36E-08	-24	1	234867300		

Table 5.1: Summary of CpG sites selected for validation as determined by Illumina Illumina Infinium 450k Beadchip assay. Δ, methylation difference between tumour and adjacent tissue; Chr, chromosome.

Final primer designs are demonstrated in Table 5.2. Melting temperatures of primers were predicted by the PyroMark Assay Design software based upon primer size and composition of nucleotides. This temperature represents the optimal annealing temperature in the PCR reaction. However, this prediction is often inaccurate and therefore it must be established experimentally.

Gene	Primer sequence	BP	Tm / °C	Amplicon size
***SPOCK*_1 (cg14650610 & cg24847829)**				542
Forward	TTATTGGTTATTGTTTAGGAAATTG	25	55	
*Reverse**	AATACTACTAAAACCCTATTCTC	23	54	
Sequencing	GAATGGGGGATTTATTTA	18	40	
***CCDC8* (cg03576469)**				330
*Forward**	TTTAGATTTTGGAAGATTGATAGG	24	55	
Reverse	AATACCCATTTCTCTACCCA	20	55	
Sequencing	AACTAAAAACTTACACAT	18	40	
***KCNQ5* (cg05447008)**				250
Forward	TGGATTGGAAAGGATGTTTAGT	22	58	
*Reverse**	CTACCCTACCTTCCAAATATTATCT	25	58	
Sequencing	ATTGGGATGGAAAGTTATA	19	44	
***BOLL* (cg13356896)**				389
*Forward**	GGTAGTTGTAGGGAAGTAGG	20	58	
Reverse	CCTTAAAATCACCTCAACCTCT	22	59	
Sequencing	AATCCCAAAACCACA	15	44	
LOC283999 (cg26034516)				387
*Forward**	GGGATTGGAGTTTAAGTTTTAG	22	55	
Reverse	CACCTCCAAAATATACCAATAATC	24	55	
Sequencing	CTAAACAATACACCCAAAAC	20	45	
***ZNF677* (cg24685755)**				400
Forward	GAAGAGGGAGTTAGAGAAGAGT	22	60	
*Reverse**	CCCTACCCTTACCCCTTAC	19	59	
Sequencing	GGGGTTTTAATTTATAGG	18	41	
***RNF220* (cg10224098)**				338
*Forward**	GTAAAGGGGAATGAGTAGAGG	21	58	
Reverse	AACCCTTCCAACTCCTAAACTA	22	58	
Sequencing	CAAACAAATATATATCCC	19	39	
Chr.1 (cg26314722)				209
Forward	ATGTAGTAGTGTTAGTAGGAGTA	23	55	
*Reverse**	CACCAATTCTTACATTTATTCACA	24	55	
Sequencing	AGTTAAAGTTAGTATAGTGT	20	42	
Chr.8 (cg09129067)				154
*Forward**	TAGGGAAAGAAGAGGATGGT	20	59	
Reverse	ACTCTCCTACCCAACCTAATACA	23	60	
Sequencing	CTTCTTACTTCTCCACTA	18	44	
Chr.14 (cg18538668)				282
*Forward**	GGTGTAGATTGGAGGATTTGT	21	59	
Reverse	TCTACTACCTCCTCTCCCT	19	58	
Sequencing	CCTCTAAATCAACCCTA	17	43	

Table 5.2: Summary of primers used for each PCR/pyrosequencing assay. BP, base pairs; Tm, melting temperature. Biotinylated primers denoted *.

5.1.2 PCR temperature optimisation

In order to determine the optimal annealing temperature of each forward/reverse primer pair, a PCR temperature gradient was performed using 100% methylated HeLA DNA as outlined in Section 4.7.3. Annealing temperatures were programmed at 50, 52, 54, 56, 58 and 60℃. PCR product was analysed by gel electrophoresis as outlined in Section 4.8.2.

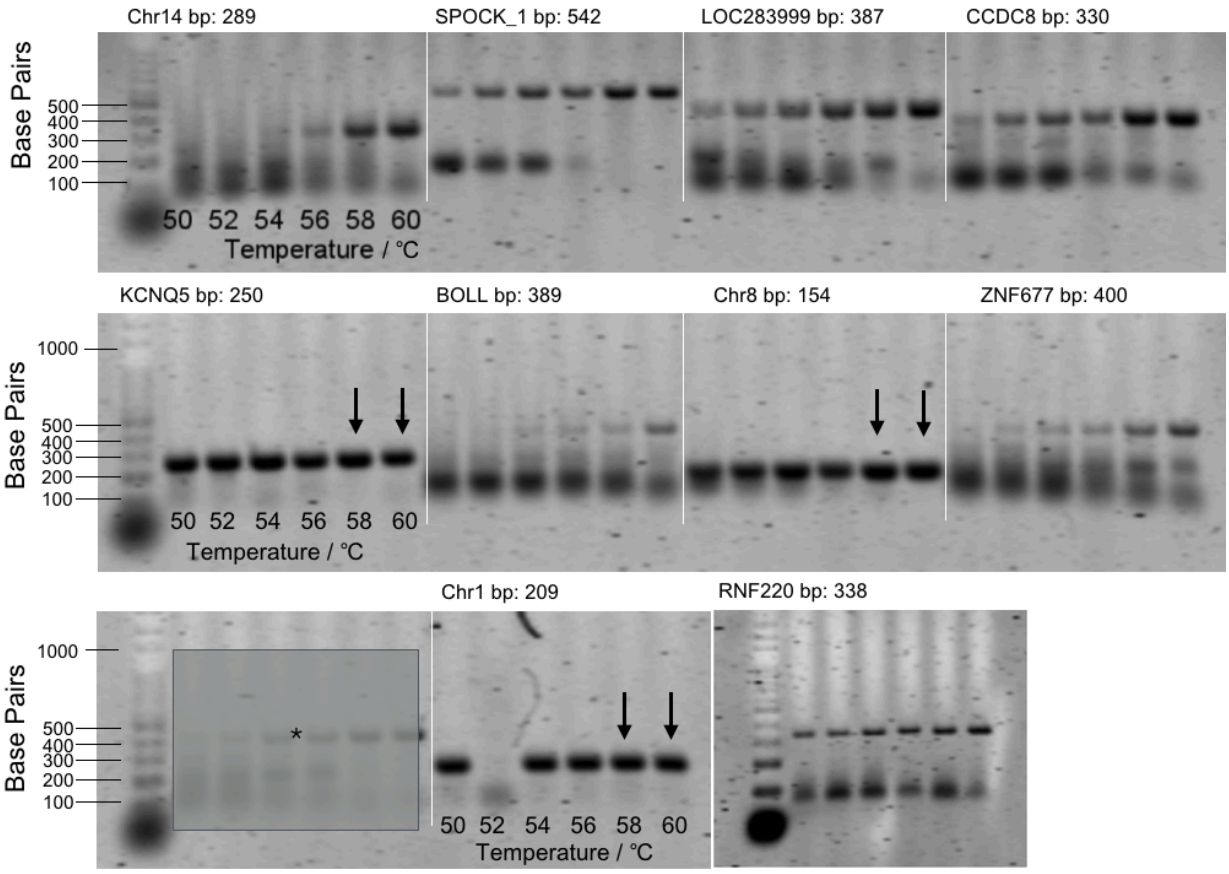


Figure 5.1: Electrophoresis gel of initial PCR temperature optimisation for all PCR primer sets. * *ZNF677* was duplicated due to a processing error and covered for clarity. *RNF220* temperature optimization gel inset. All PCR reactions were most effective at 58-60°C. The two optimum annealing temperatures selected for pyrosequencing are highlighted with black arrows. Those without arrows were assessed beyond 60°C.

All primer sets successfully resulted in amplicon formation of the correct molecular weight (Figure 5.1). Chr14, *SPOCK*_1, LOC283999, *CCDC8*, *BOLL, RNF220 and ZNF677* each displayed maximum intensity of PCR product at 60℃. To determine whether the efficiency of the reaction continued to increase beyond 60℃, additional PCR reactions with annealing temperature gradients ranging from 56-66°C were performed. All other reaction conditions were maintained from the initial gradient.

Increased efficacy above 60°C was noted in Chr14, *LOC28399, CCDC8, BOLL and ZNF677* (Figure 5.2). The two temperatures with maximum PCR product in each primer set were selected for pyrosequencing (Table 5.3).

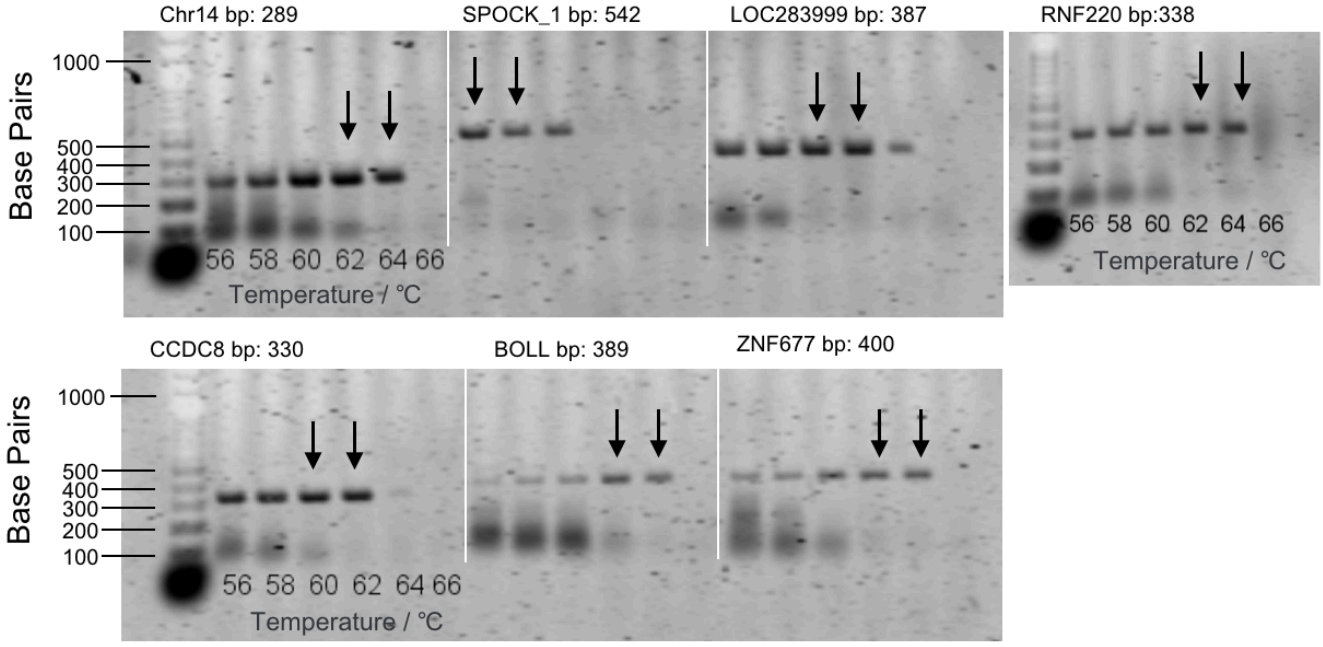


Figure 5.2: Electrophoresis gel of further PCR annealing optimization from 56-66°C. Optimum annealing temperatures selected for pyrosequencing are highlighted by black arrows. A 100bp ladder is used to quantify amplicon size.

Gene	Optimum annealing temperatures /°C
*SPOCK*_1 (cg14650610 & cg24847829)	58 & 60
CCDC8 (cg03576469)	60 & 62
KCNQ5 (cg05447008)	58 & 60
BOLL (cg13356896)	62 & 64
LOC283999 (cg26034516)	60 & 62
ZNF677 (cg24685755)	62 & 64
RNF220 (cg10224098)	62 & 64
Chr1 (cg26314722)	58 & 60
Chr8 (cg09129067)	58 & 60
Chr14 (cg18538668)	62 & 64

Table 5.3: Two optimum annealing temperatures for each primer set as determined by gel electrophoresis for further analysis by pyrosequencing.

5.2 Pyrosequencing optimisation

In order to determine which PCR annealing temperature produced the optimum product for pyrosequencing, the two PCR products highlighted in Table 5.3 (Section 5.1.2) were selected from each assay. Pyrosequencing assays were designed to assess each DMP as described in Section 4.9.2. Initial pyrosequencing was performed using 15µl of PCR product as outlined in 4.9.4.

The optimum PCR annealing temperatures for pyrosequencing are summarised in Table 5.5. *KCNQ5* and Chr8 displayed no considerable difference in assay performance between temperatures analysed. In these cases, the higher temperature was selected in order to reduce primer dimer formation.

BOLL and *RNF220* demonstrated evidence of secondary annealing around their respective target CpGs and were therefore excluded from further experiments. All other assays successfully produced peaks in the nucleotide positions expected.

Additional changes were made to the volume of PCR product analysed and nucleotide dispensation orders as to increase pyrosequencing signal and to account for prolonged refractory times of repeated dNTP sequences respectively. Modified assays were repeated under new conditions in duplicate and further improvements were made as necessary. Final dispensation orders and the volume of PCR product required for pyrosequencing are summarised in Table 5.4.

The ability of assays to accurately assess DNA methylation at the target CpG site using 100% methylated HeLa DNA is shown in Figure 5.3. Of note, some assays such as *KCNQ5* and Chr1 appreciably underestimate the value of DNA methylation at these sites. This underestimation may also persist when analysing DNA obtained from our CRC cohorts. The correlation between Illumina 450k data and data obtained from pyrosequencing is therefore more important than matching absolute values.

DMP	Volume of PCR product / µl	Final dispensation sequence
ZNF677	25	TGTCGATTAATCGATCATGTCGATGTCGTGATGTAGTCATGTCGATGTCGTCA
KCNQ5	15	TGTCGTCAGTGAGTGATGTCATGTCGAGATGTGTAGTTCATGTCGATGTCATGTCTATGTCGTTAGTCGTG
LOC283999	25	CAGATCAGACAAACATCGATAGACGACATCAGACATGAAT
*SPOCK*_1	25	ATCGTATCGAGAGTATGTCGCTGATCGTATCGAGTAGTCGTATATCAGGTATCAGTTCGT
CCDC8	25	CTACAGATACATAATACGACTAGCTCAAACATAGCTATCGATACACTATCGACTACGACTAA
Chr1	15	ATTTGTCATGAGTATATTTATCATGATAGTGTTATCGTTA
Chr8	15	CTCTCCAGACTCTACAGATCGACTGACACATACGATACGATCGATCT
Chr14	15	CATCACTATGATACGACTACCGATAACTGACTACTCAACTC

Table 5.4: Final nucleotide dispensation orders and volume of PCR product required for pyrosequencing assays.

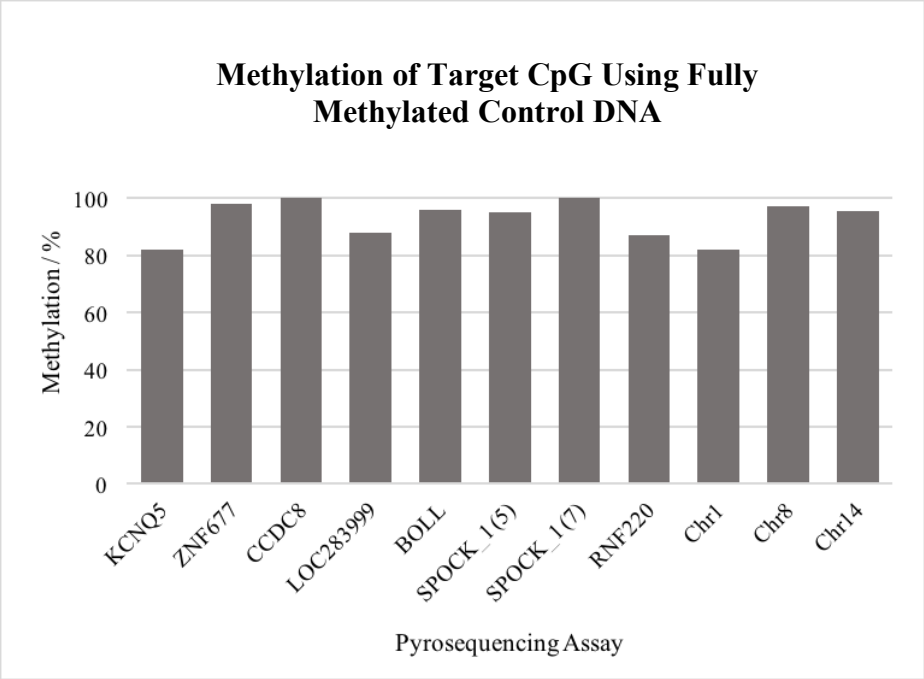


Figure 5.3: Methylation value of target CpG as calculated by each pyrosequencing assay using 100% methylated HeLa DNA as control.

5.3 Additional assay optimisation

Significant primer-dimer formation was noted in the gel images of LOC283999, *BOLL, ZNF677* and Chr14 (Figures 5.1 and 5.2). As secondary sequences can produce background signal during pyrosequencing, further optimisation was conducted to reduce the amount of primer-dimer formed.

PCR was therefore conducted using 0.4, 0.8, 1.2 and 1.5µl of 10µM forward and reverse primer for each 30µl reaction. Gel electrophoresis demonstrated reduced primer-dimer formation and relatively increased desired product at lower primer concentrations (Figure 5.4). Subsequent pyrosequencing demonstrated that 0.8µl of forward and reverse LOC283999 primer solution resulted in reduced background noise and peak height deviation, whilst maintaining sufficient peak height for successful analysis. Adjusted primer volume had no impact on the background noise or methylation values of other assays assessed, however did result in reduced peak height.

Optimised PCR conditions for pyrosequencing assays are summarised in Table 5.5.

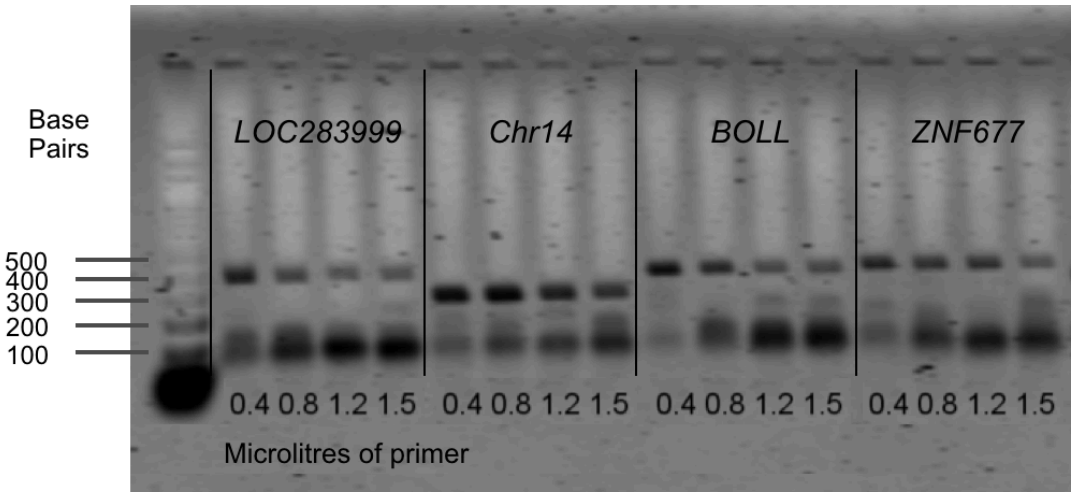


Figure 5.4: Primer volume optimization gel demonstrating reduced primer-dimer and increased product at lower primer concentrations.

	ZNF677	*KCNQ5*	*SPOCK_1*	LOC283999	*CCDC8*	Chr1	Chr8	Chr14
10x B1 Buffer / µl	3.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0
MgCl/ µl	2.0	1.3	2.0	2.0	2.0	1.3	1.3	1.3
dNTP/ µl	0.3	0.2	0.3	0.3	0.3	0.2	0.2	0.2
F/R Primer/ µl	1.5	1.0	1.5	0.8	1.5	1.0	1.0	1.0
Hotstart Taq/ µl	0.3	0.2	0.3	0.3	0.3	0.2	0.2	0.2
H20/ µl	19.9	13.3	18.9	20.6	19.9	13.3	13.3	12.3
DNA/ µl	3.0	2.0	4.0	3.0	3.0	2.0	2.0	3.0
Total	30.0	20.0	30.0	30.0	30.0	20.0	20.0	20.0
Annealing temperature / °C	62.0	60.0	58.0	60.0	62.0	60.0	60.0	62.0

Table 5.5: Summary of optimum PCR conditions for downstream bisulphite pyrosequencing assays.

5.4 Optimisation of the LUMA assay

Global DNA methylation was determined using the LUMA assay as outlined in Section 4.10. As discussed, multiple modifications have been described since the method was initially proposed. Modified nucleotide dispensation orders and the use of the MunI restriction enzyme in place of EcoRI, aim to fill non-specific overhangs and reduce STAR activity respectively. In order to determine the effect of these modifications, variations of dispensation orders and restriction enzyme were compared. The ability to include technical replicates with 150ng of genomic DNA was also assessed. DNA obtained from H37T was used as control.

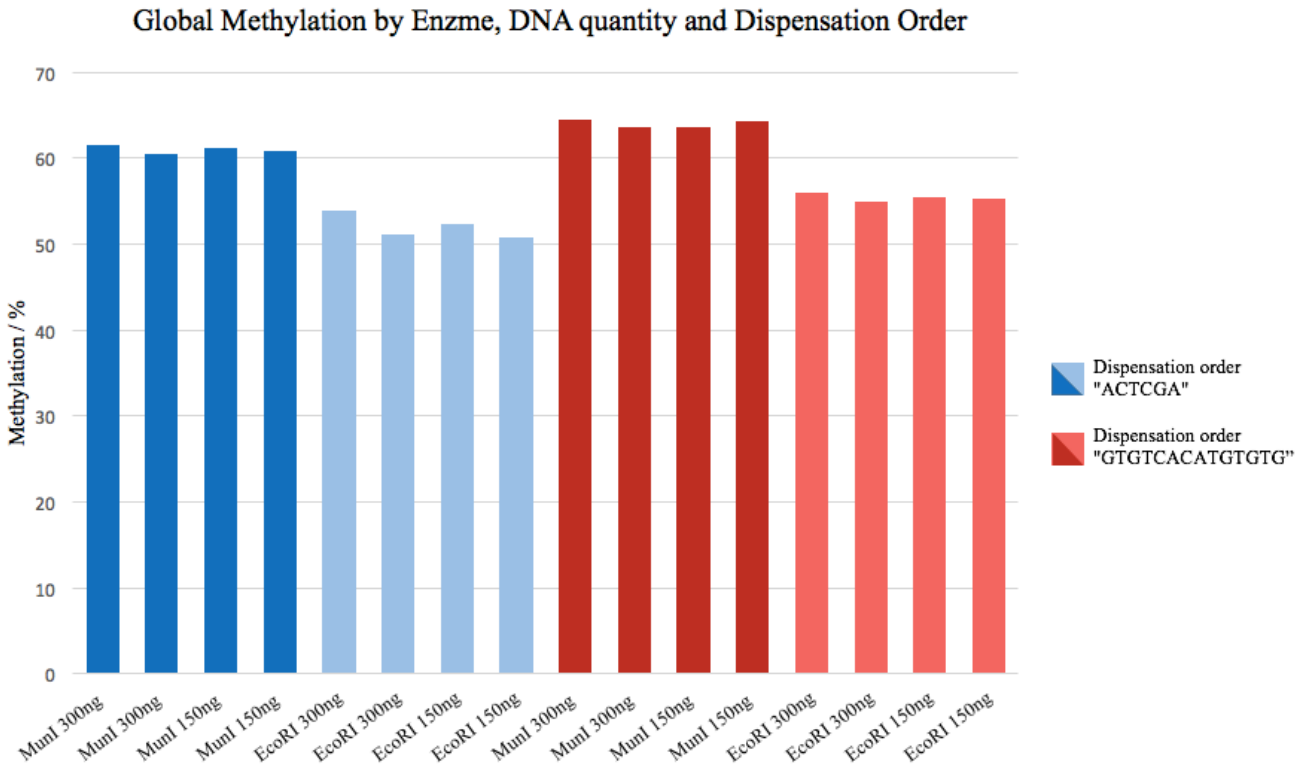


Figure 5.5: Methylation values determined by the LUMA assay under varying conditions. Restriction enzymes MunI and EcoRI. The GTGTCACATGTGTG dispensation order and the use of the MunI restriction enzyme in place of EcoRI, aim to fill non-specific overhangs and reduce STAR activity respectively.

The assay upheld reproducibility between all duplicated conditions (Figure 5.5). Little variation was seen between DNA methylation levels determined using 150ng and 300ng of DNA. Use of the restriction enzyme MunI and the

modified dispensation order (GTGTCACATGTGTG) each resulted in higher levels of global DNA methylation as expected.

5.5 Identification of eligible samples for LUMA assay

The LUMA assay works on the assumption that all fragmentation occurs as result of restriction enzymes. Therefore, the integrity of DNA used must be determined prior to analysis.

The analysis was performed using DNA obtained from RIST samples. Sufficient DNA (600ng) was available to perform LUMA analysis of 11 of 14 matched sample pairs. The quality of DNA was assessed by electrophoresis using a 0.87% ethidium bromide stained gel as described in 4.8.3 (Figure 5.6).

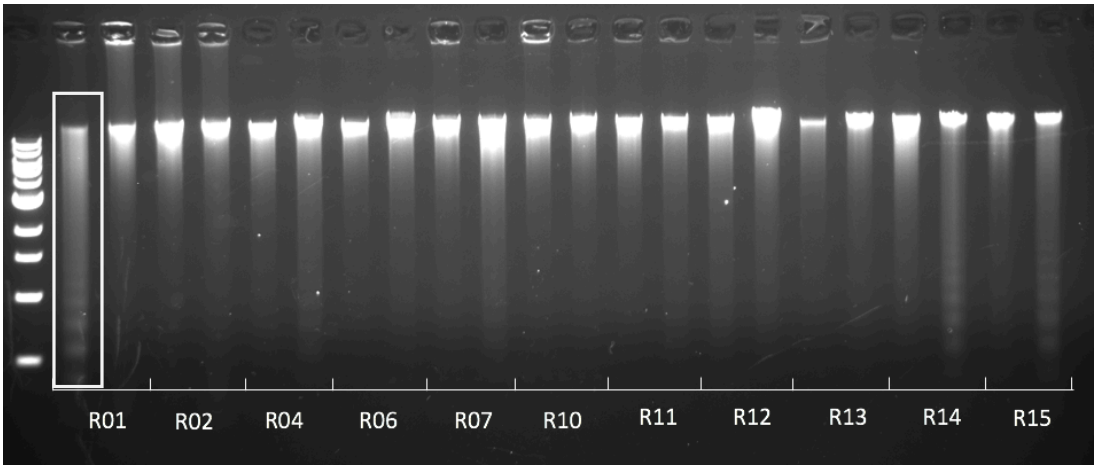


Figure 5.6: Ethidium bromide stained gel of DNA obtained from RIST samples. The smeared band of DNA sample R01N represents degradation/fragmentation (white box).

DNA from sample R01N was highly fragmented and therefore excluded from the analysis. Some fragmentation was also seen in R04T, R07T, R014T and R015T. However due to the relatively large amount of high molecular weight product seen in these samples, the significance of this fragmentation was deemed low. These samples were therefore included in LUMA analysis.

6 Results: Application of Optimised Assays to Discovery Samples

6.1 Application of pyrosequencing assays to discovery set

DMPs identified in our preliminary studies were identified by an EWAS using the Illumina 450k assay. In order to technically validate these findings, assays developed in the previous Section were applied to the same DNA samples (RIST cohort) as the preliminary work.

The concentration of DNA obtained from RIST samples (15 tumour and 15 matched adjacent mucosa) was quantified by photospectrometry and bisulphite treated as described in Sections 4.4 and 4.6. PCR and pyrosequencing assays were applied to this bisulphite converted DNA. Failed analyses of individual samples were repeated alongside their corresponding tumour/adjacent samples.

ZNF677, LOC283999 and all hypomethylated DMPs were successfully applied to all DNA samples. The forward and reverse primers of the SPOCK_1 assay consistently failed to amplify DNA from tumour sample six (T006), although adjacent/normal sample 6 (N006) amplified successfully, indicating a potential mutation within one of the primer binding sites. With regard to *KCNQ5, SPOCK_1* and *CCDC8*, several DNA samples failed to amplify in pairs (i.e. both tumour and adjacent mucosa, Figure 6.1).

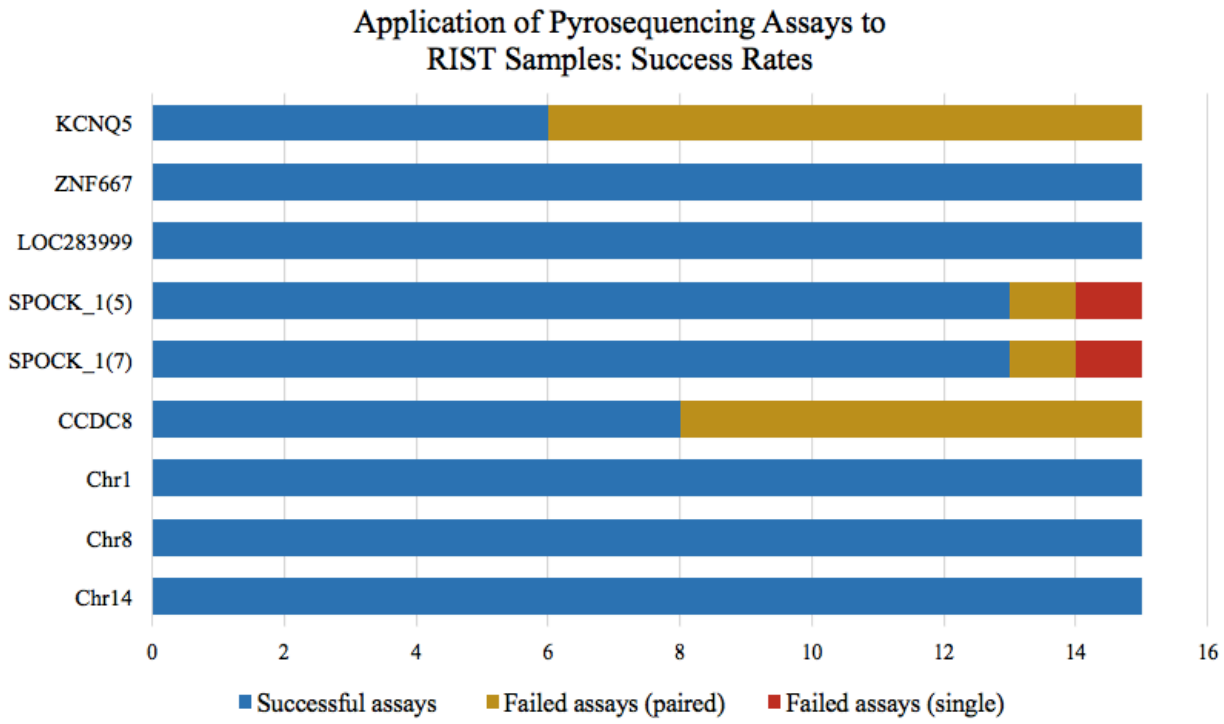


Figure 6.1: Graph to show the number of successfully amplified RIST samples for each primer set. Unsuccessful assays are subdivided into those that failed in pairs (tumour and adjacent tissue) and those in which only a single sample failed to amplify.

6.2 Single nucleotide polymorphisms

Failure to amplify both tumour and normal tissue DNA obtained from the same patient was observed in nine, seven and one cases of the *KCNQ5, CCDC8 and SPOCK_1* assays respectively (Figure 6.2). We hypothesised that this paired amplification failure could result from the presence of SNPs within the forward and/or reverse primer binding sites of each assay, which in turn could interfere with PCR amplification.

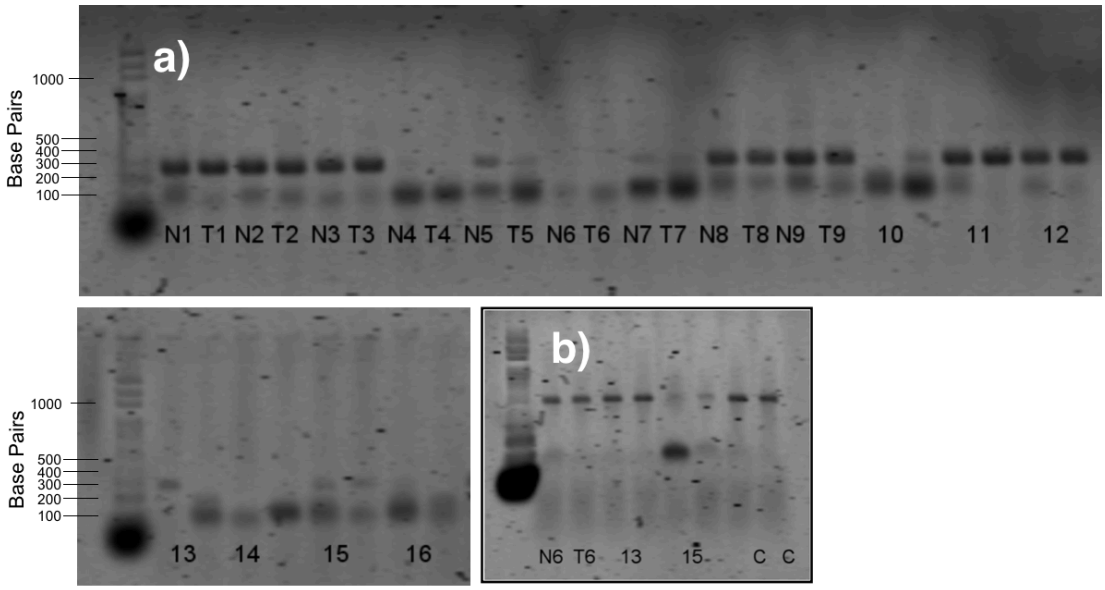


Figure 6.2: Electrophoresis gel of a) *KCNQ5* and b) *SPOCK*_1 amplification of RIST samples. Both tumour and non-tumour DNA obtained from patients R04, R05, R06, R07, R10, R13, R14, R15 and R16 failed to amplify *KCNQ5*. Paired amplification failure of R15 was also observed in R15. T, tumour sample DNA; N, non-tumour/adjacent tissue sample DNA.

The UCSC genome browser was used to identify common SNPs within these regions. (148) Two common SNPs were identified within the biotinylated reverse primer of *KCNQ5*. To avoid the 5' C/A SNP, the primer binding site was moved in a 3' direction and an A/G substitution of the 3' SNP was made (Figure 6.3). The modified primer sequence was ACAACCCTACCCTACCCTC. One uncommon G/A SNP was also identified

within the forward primer, however this was reported to occur with low frequency (0.09%) and was therefore overlooked.

Figure 6.3: SNPs identified within the primer sequence of the reverse *KCNQ5* primer (highlighted in red). BSC, bisulphite converted.

One common SNP (A/G) was identified within the reverse primer of *SPOCK*_1 (Figure 6.4). Although three SNPs were identified within the forward primer, these occurred at frequencies less than 1%. Therefore, no modifications were made to the forward primer. The revised reverse primer sequence was AATACTATTAAAACCCTATTCTC.

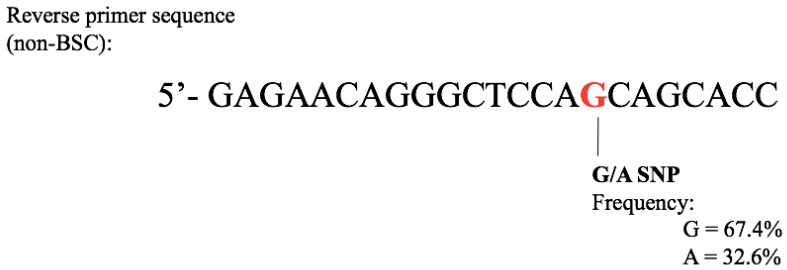


Figure 6.4: SNP identified within the primer sequence of the reverse SPOCK_1 primer (highlighted in red). BSC, bisulphite converted.

Only one C/T SNP was identified within the reverse primer of *CCDC8*. However these were relatively uncommon, occurring in 98.2% (T) and 1.8% (C) of the population. Primer modification was not attempted for this DMP.

Modified primers were optimised using 100% methylated HeLA DNA as previously described. The optimum PCR conditions of these assays is shown in Table 6.1. The pyrosequencing dispensation orders were unchanged.

	***SPOCK*_1 (SNP)**	***KCNQ5* (SNP)**
10x B1 Buffer	3	2
MgCl	2	1.3
dNTP	0.3	0.2
F/R Primer	1.5	1
Hotstart Taq	0.3	0.2
H20	19	13.3
DNA	4	2
Total	30	20
Temperature / °C	56	62

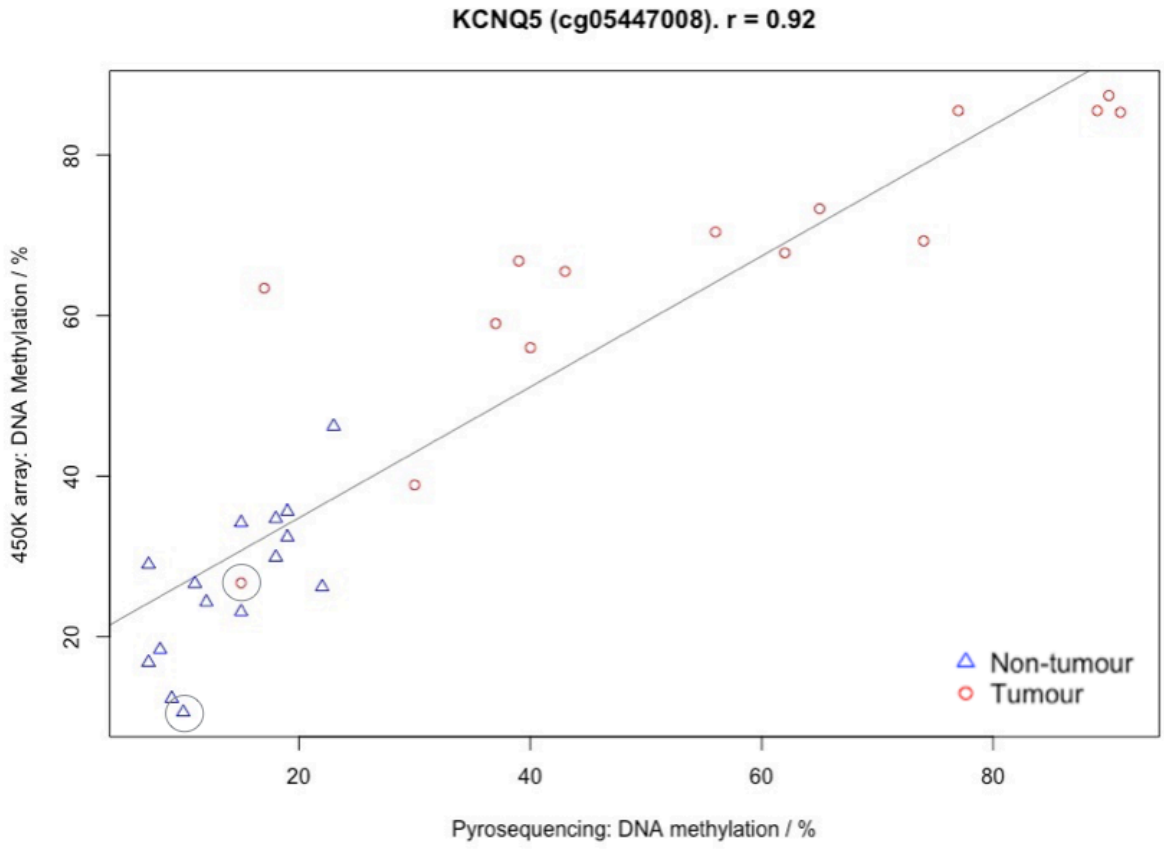
Table 6.1: Optimum PCR conditions of modified primers.

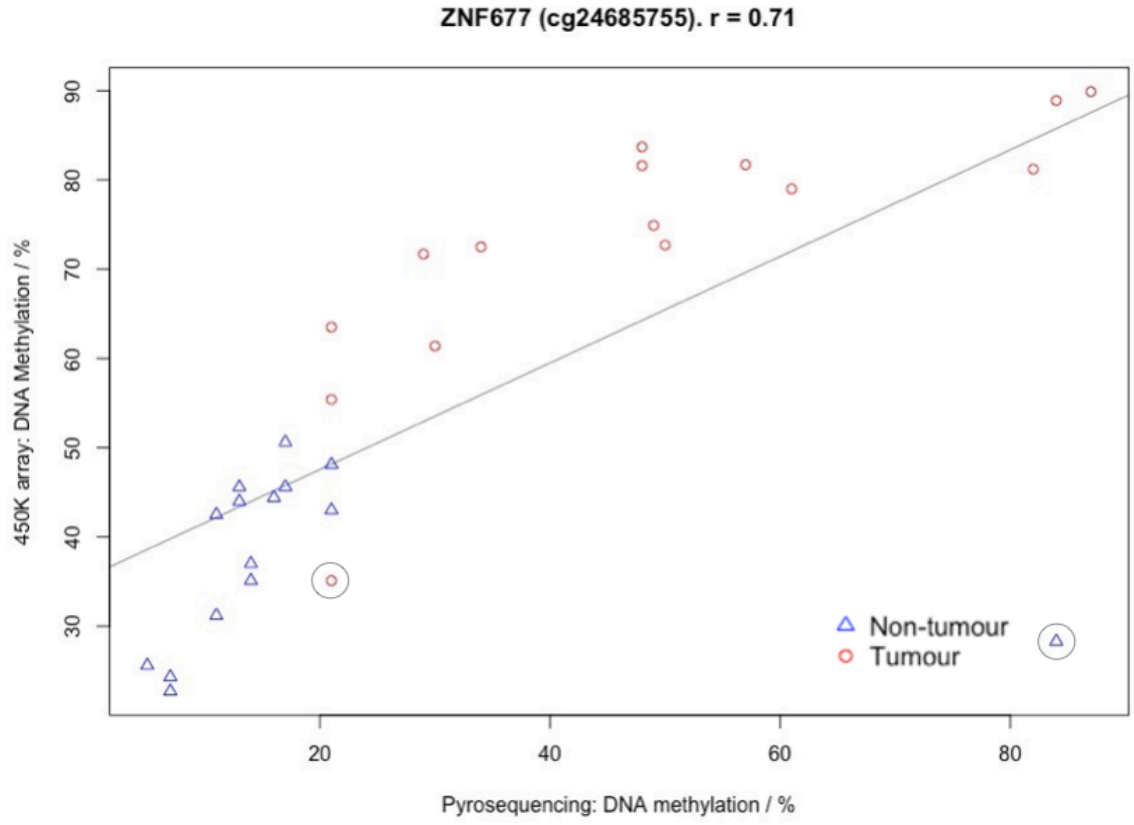
Application of these primers to previously failed samples resulted in successful amplification and pyrosequencing in all cases.

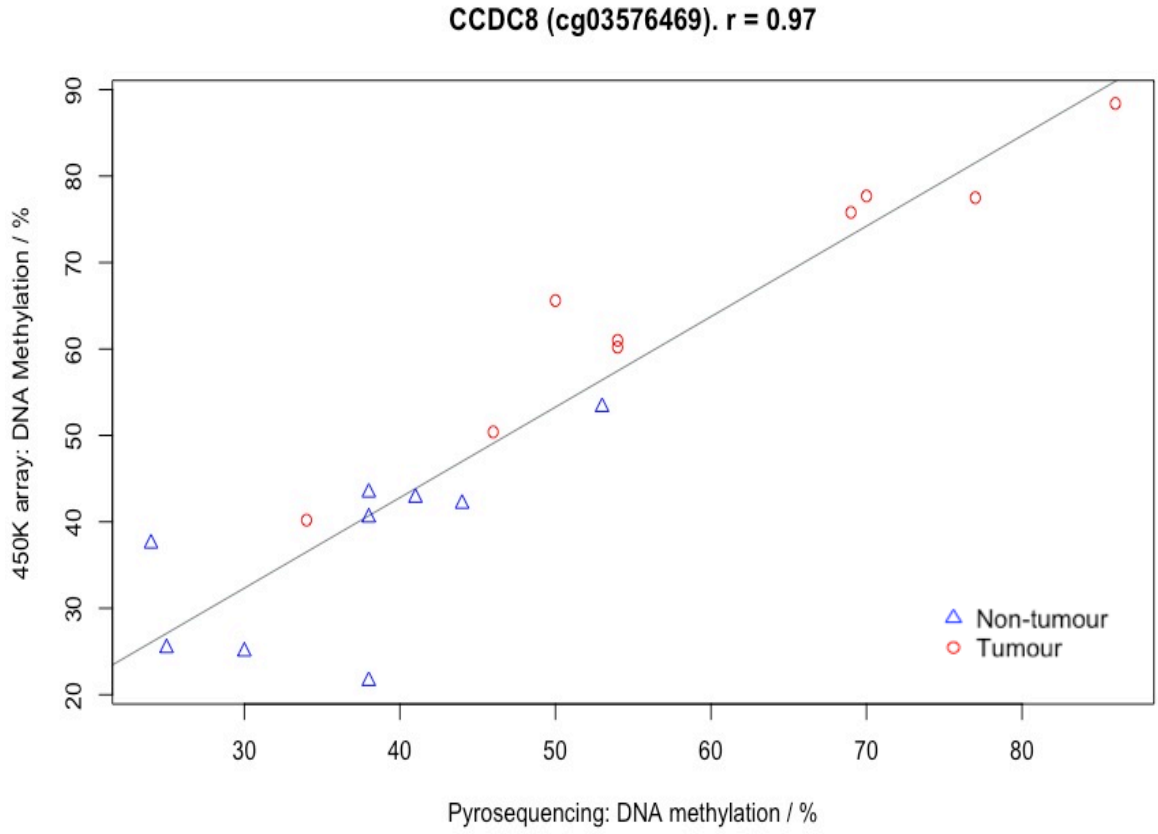
6.3 Correlation of pyrosequencing data with 450k

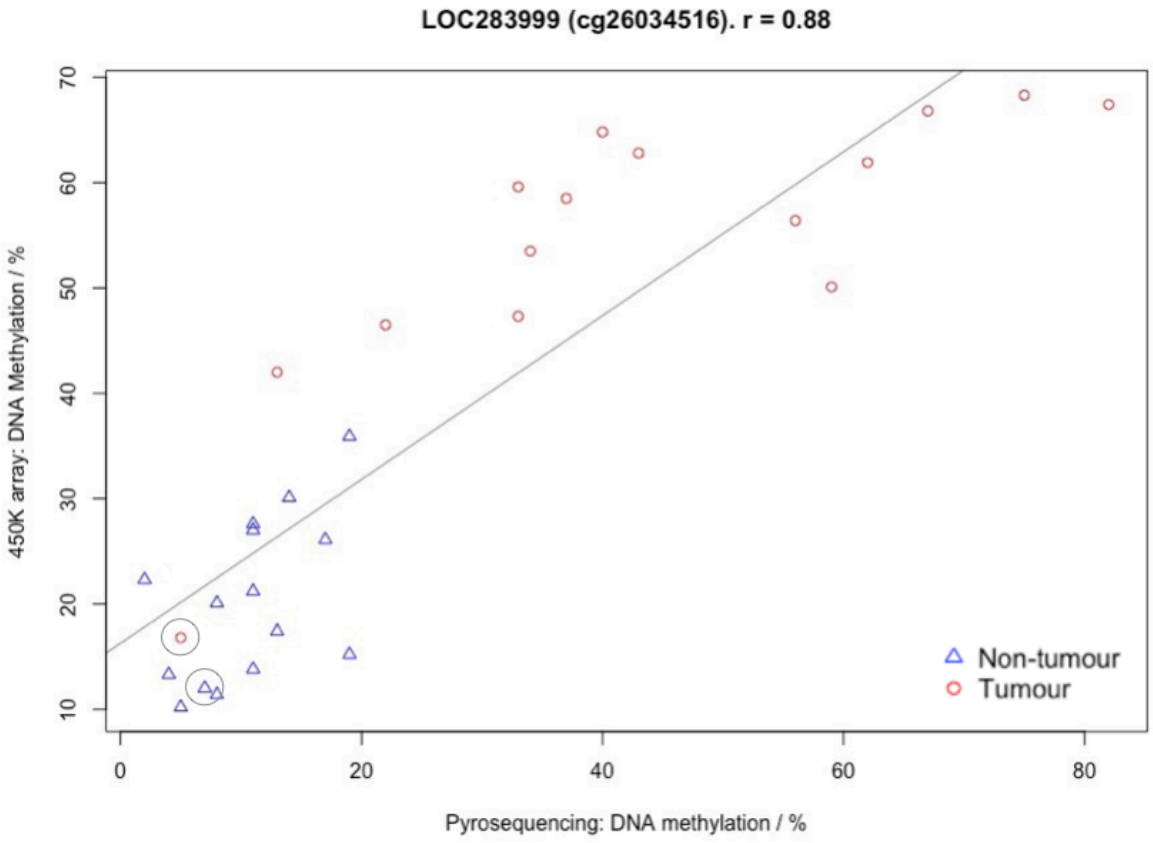
The correlation between methylation values determined by genome-wide analysis and pyrosequencing was assessed by calculating the Pearson's correlation coefficient (r, Figure 6.5).

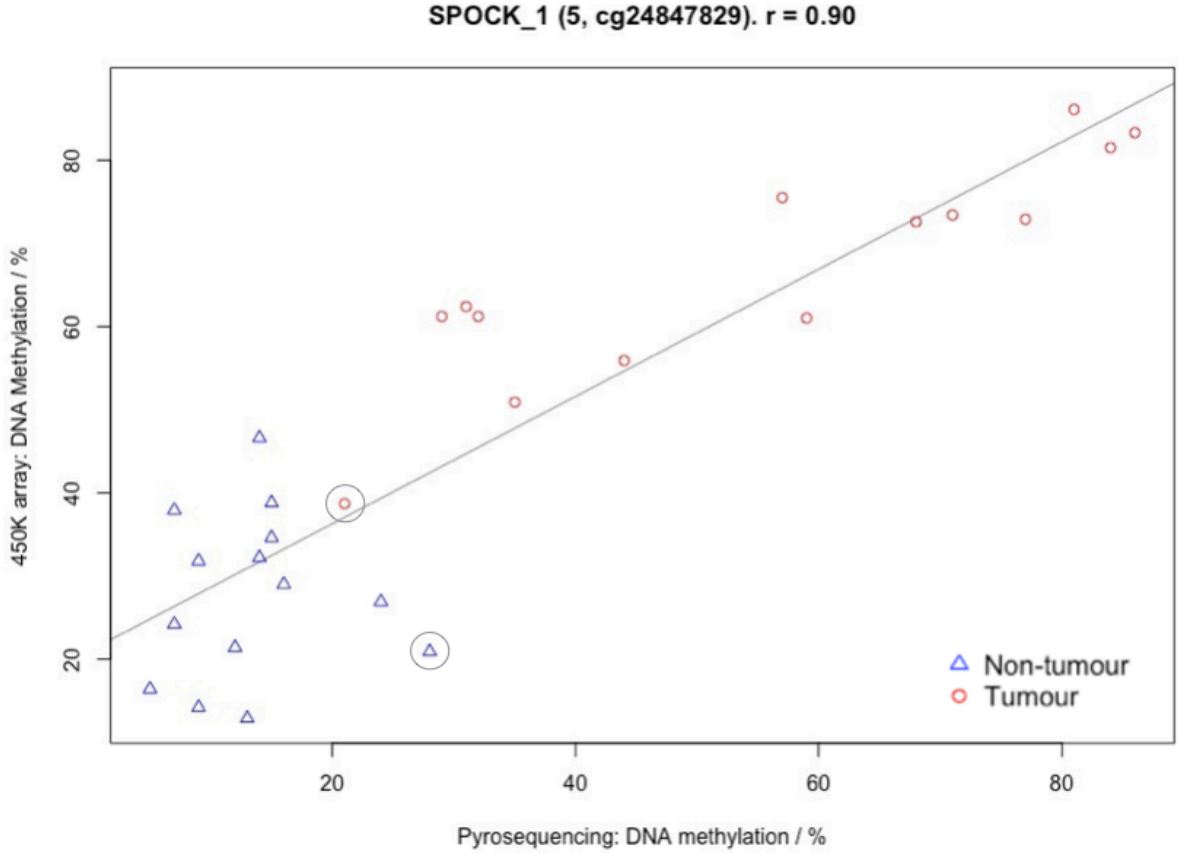
Strong correlations were observed between assessment methods of all DMPs analysed. However, considerable discordance was observed in the methylation values of DNA samples T003 and N003 (circled Figure 6.5) in assays assessing *ZNF677*, Chr1 and Chr14. T003 also demonstrated methylation values consistent with those of non-tumour tissue at all other loci. These anomalies persisted following multiple repeats. Given the strong concordance of all other samples, this was likely caused by a sample error and therefore N003 and T003 were excluded from further analyses. Exclusion of these samples resulted in fewer outlying data points and strengthened the correlation between assessment methods (Figure 6.6).

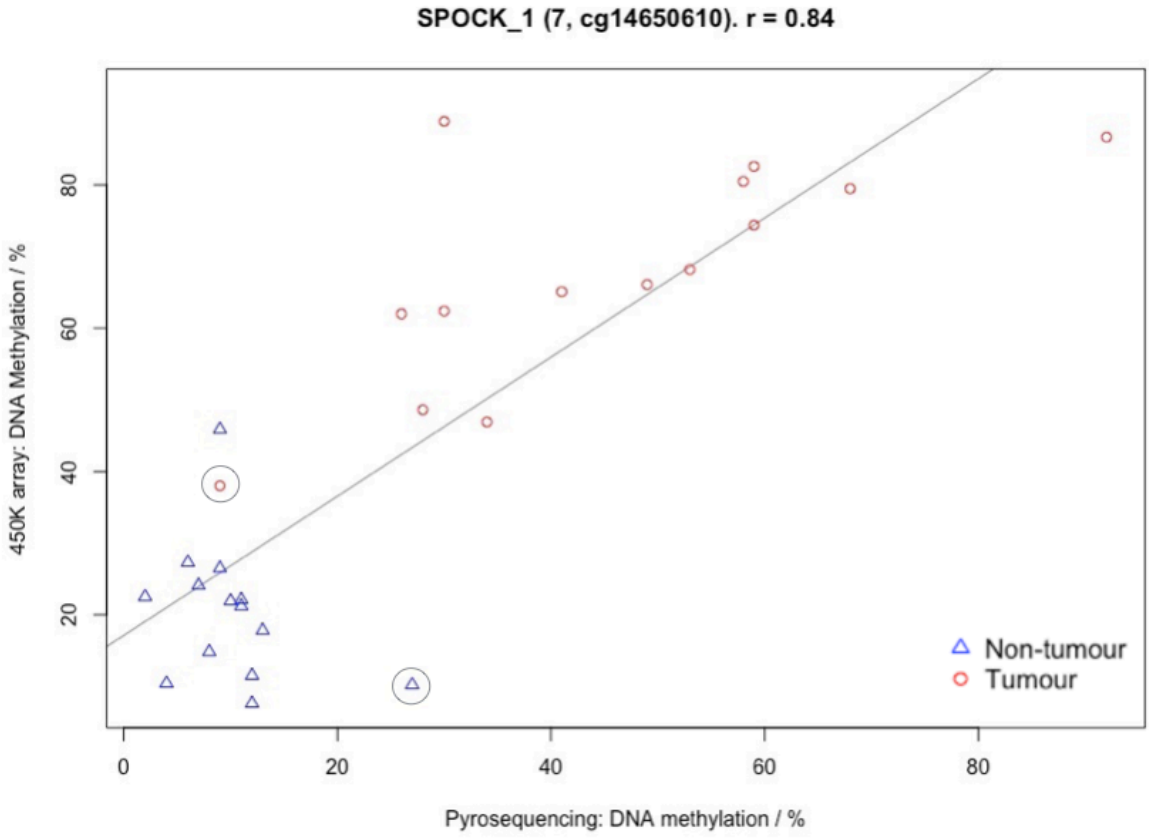


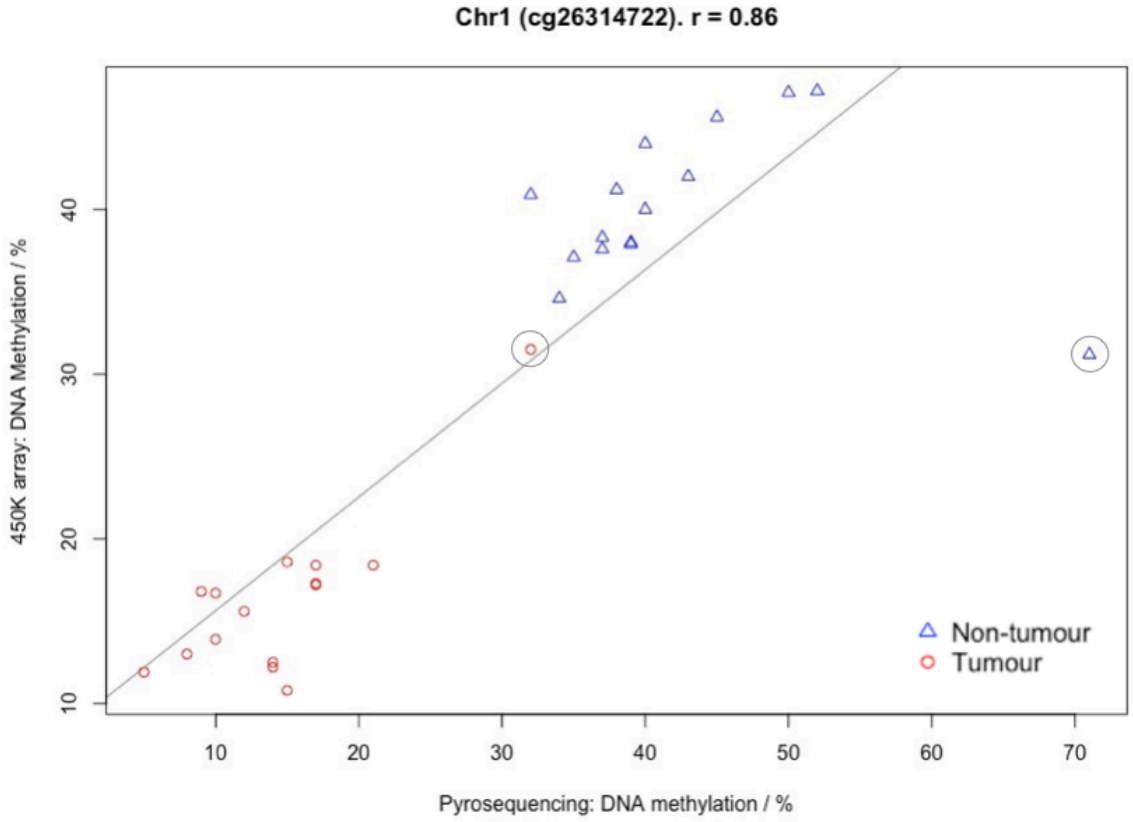

ZNF677 (cg24685755). r = 0.71
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour

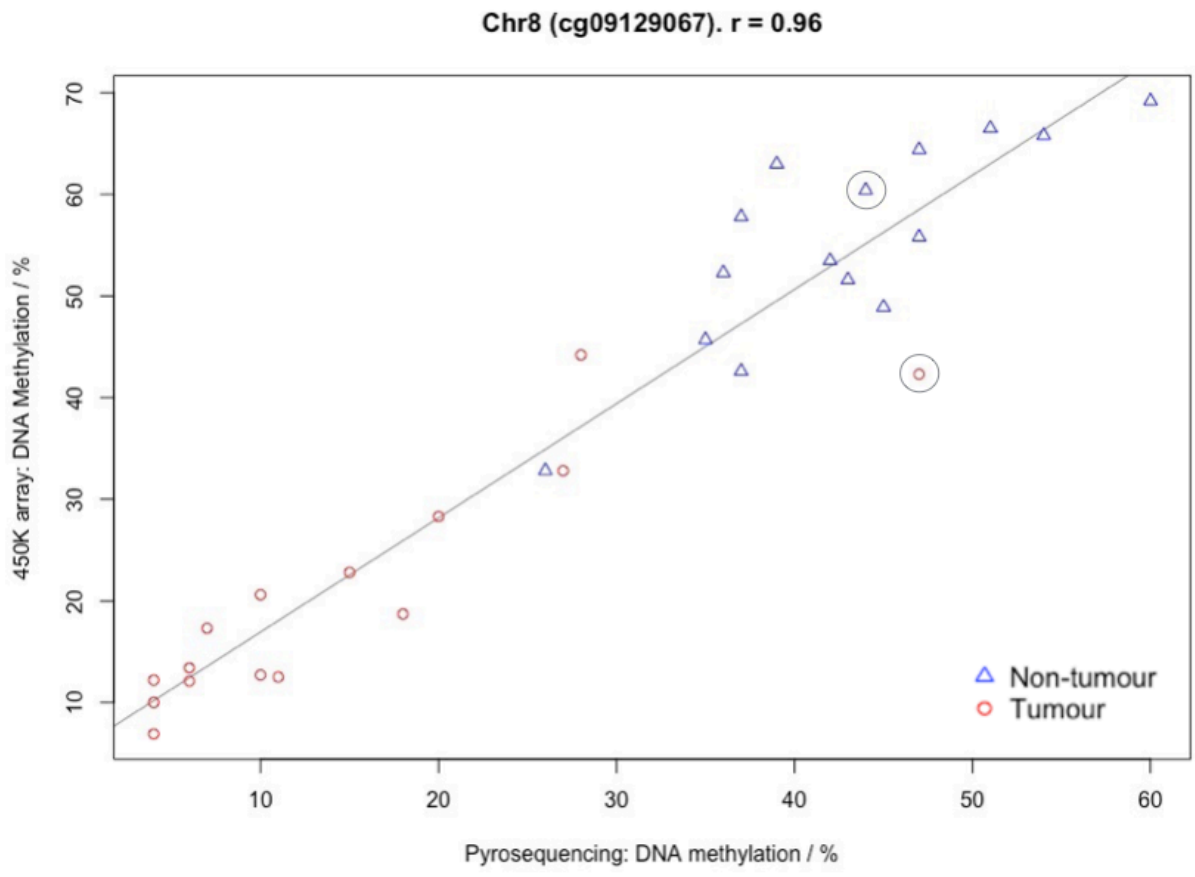

CCDC8 (cg03576469). r = 0.97
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour


LOC283999 (cg26034516). r = 0.88
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour


SPOCK_1 (5, cg24847829). r = 0.90
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour


SPOCK_1 (7, cg14650610). r = 0.84
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour
0
20
40
60
80


Chr1 (cg26314722). r = 0.86
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour
10
20
30
40
50
60
70


Chr8 (cg09129067). r = 0.96
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour

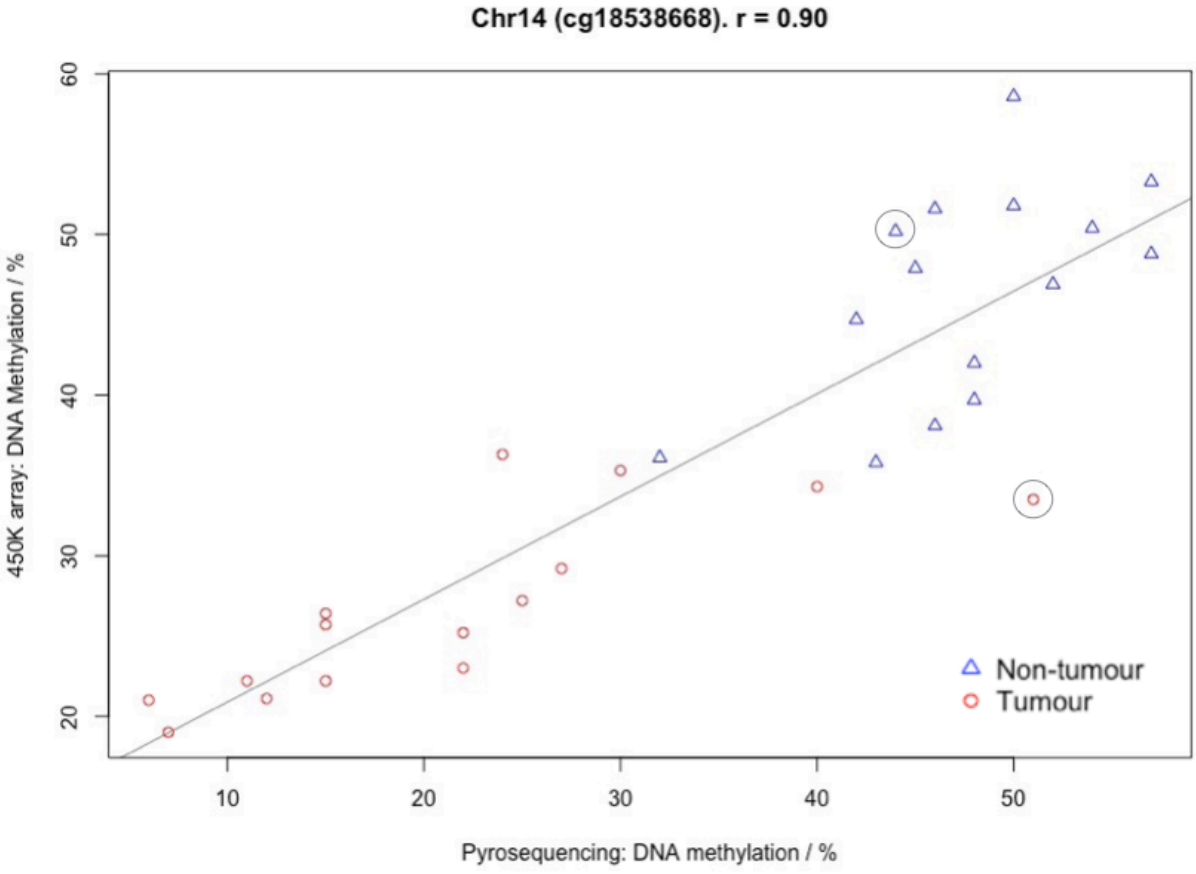
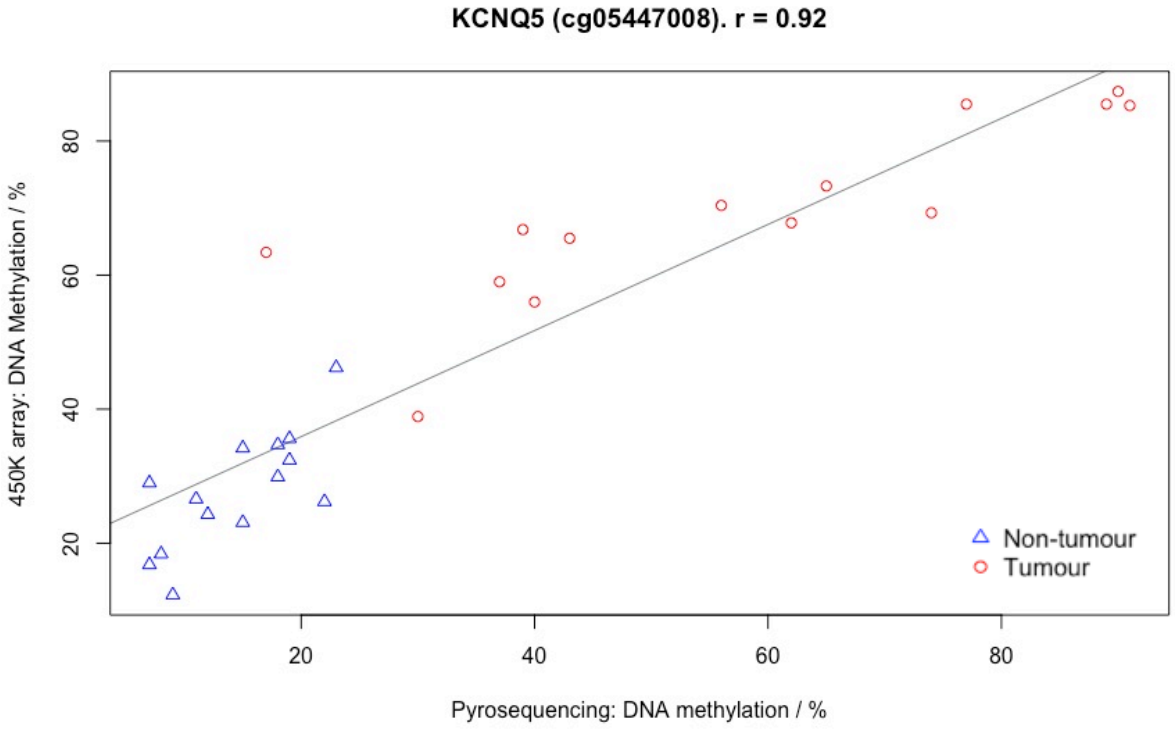
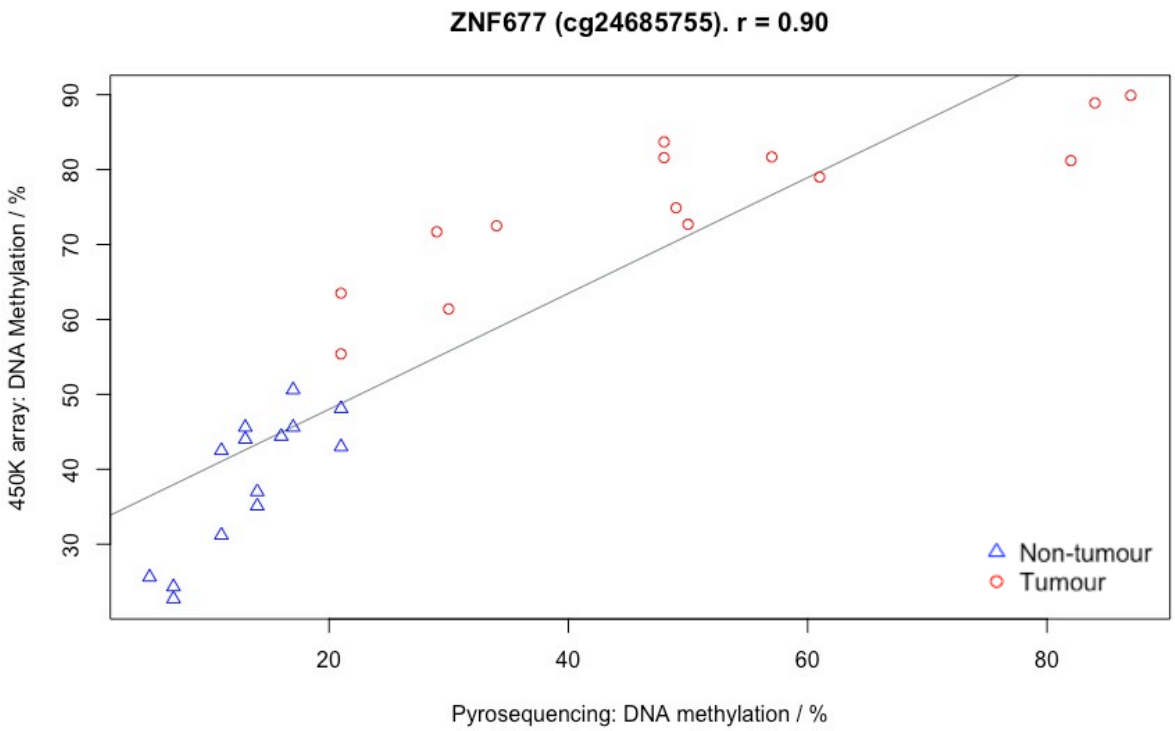
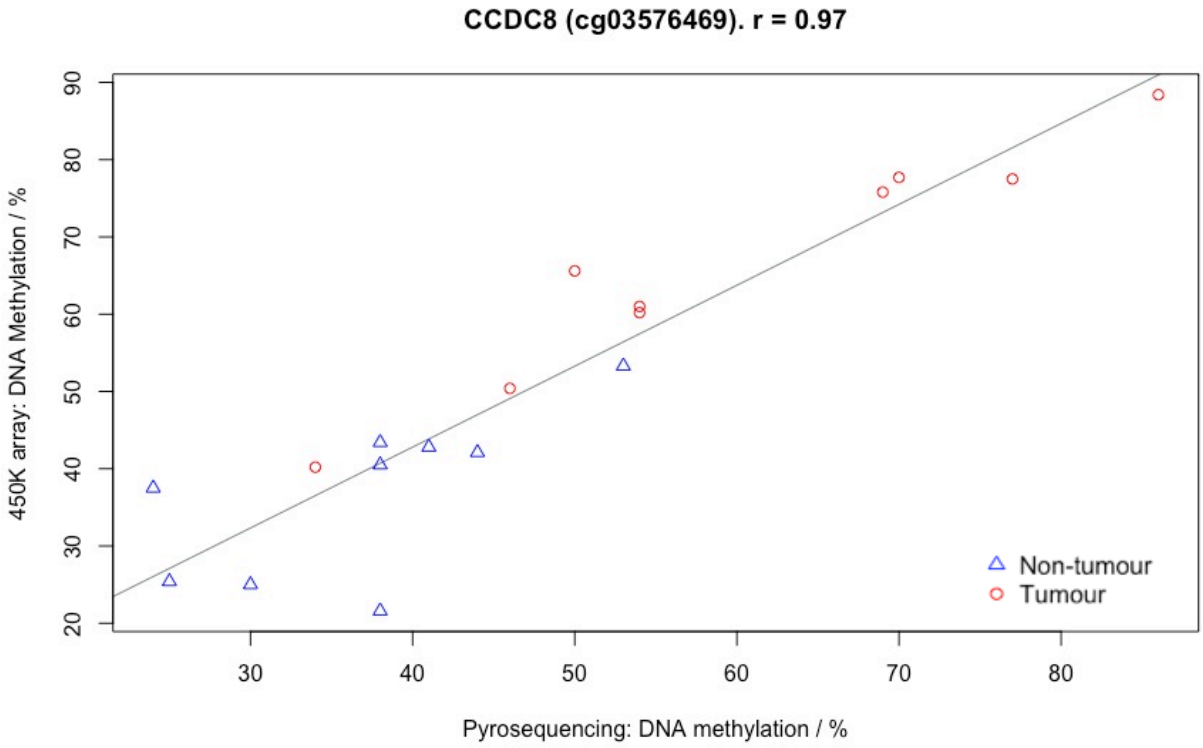
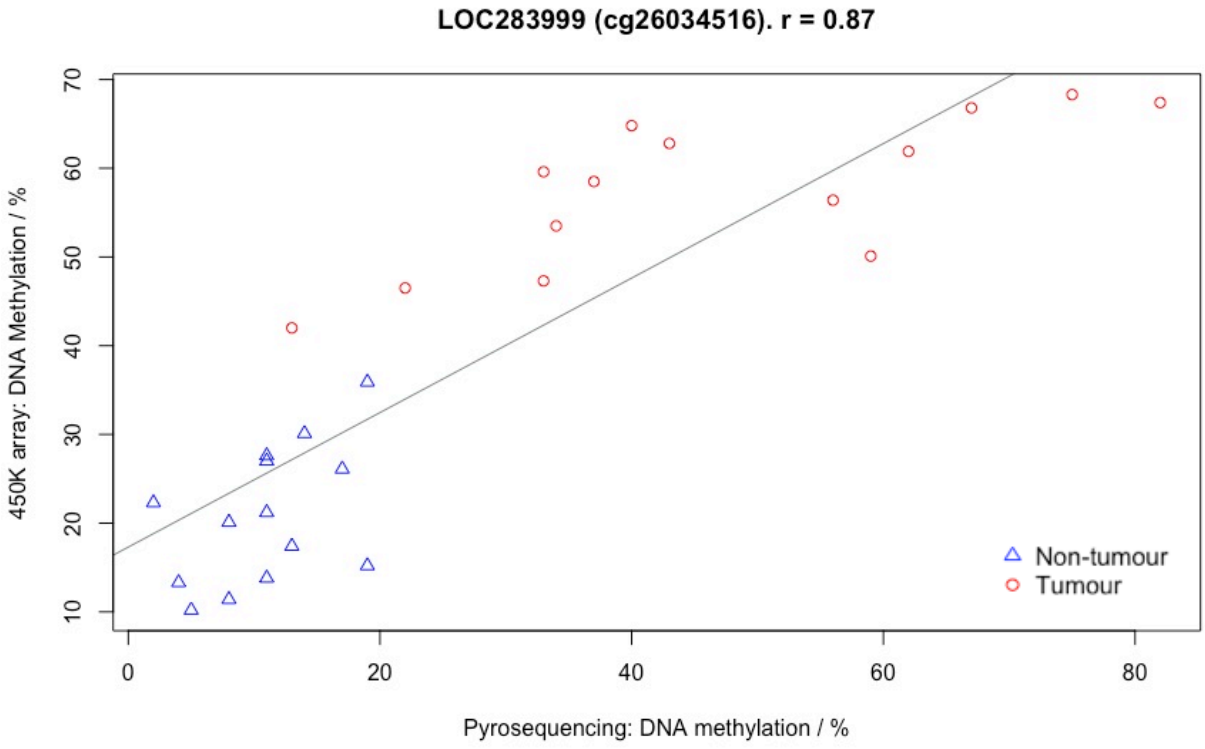

Chr14 (cg18538668). r = 0.90
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour

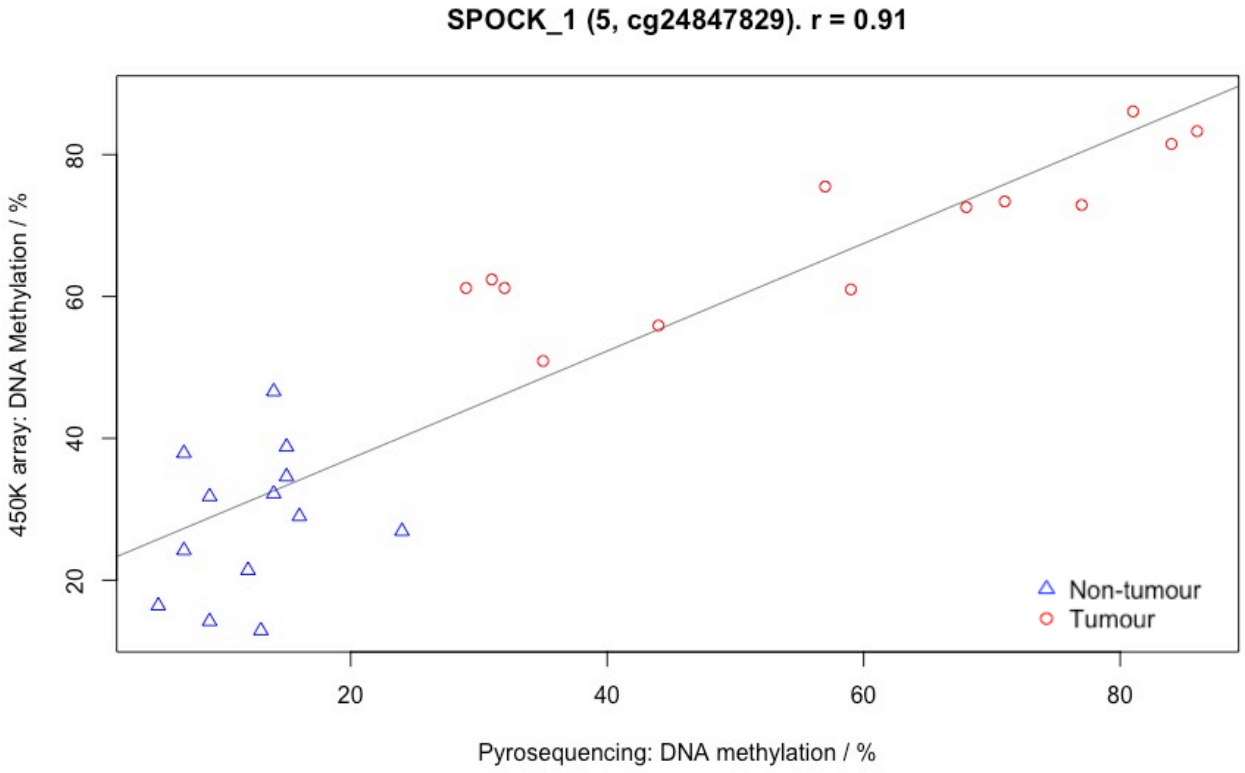
Figure 6.5: Scatterplots to demonstrate the correlation between methylation values of DMPs assessed by pyrosequencing and the Illumina 450k genome wide assay. Values corresponding to RIST sample R03 are circled. These samples displayed considerable discordance between DNA methylation levels obtained from pyrosequencing and the 450K array despite multiple repeats. r, Pearson's correlation coefficient.

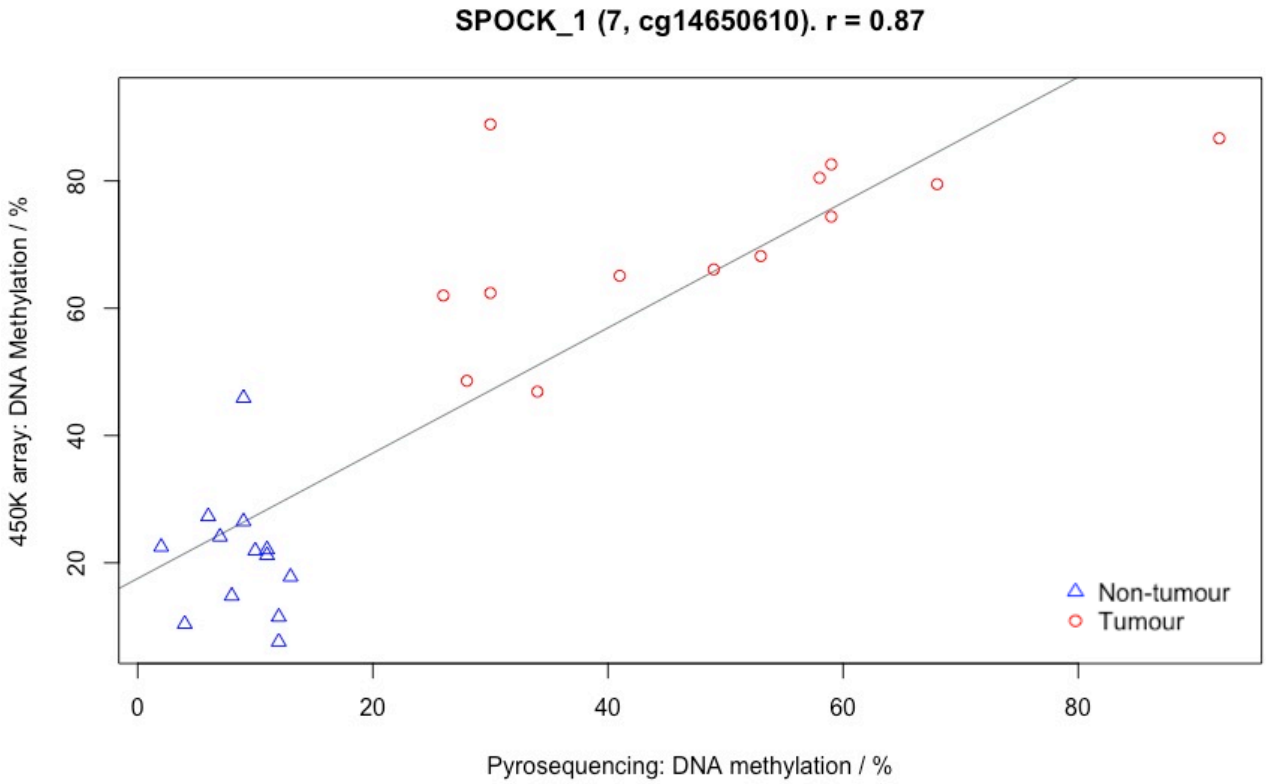

KCNQ5 (cg05447008). r = 0.92
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour
20
40
60
80

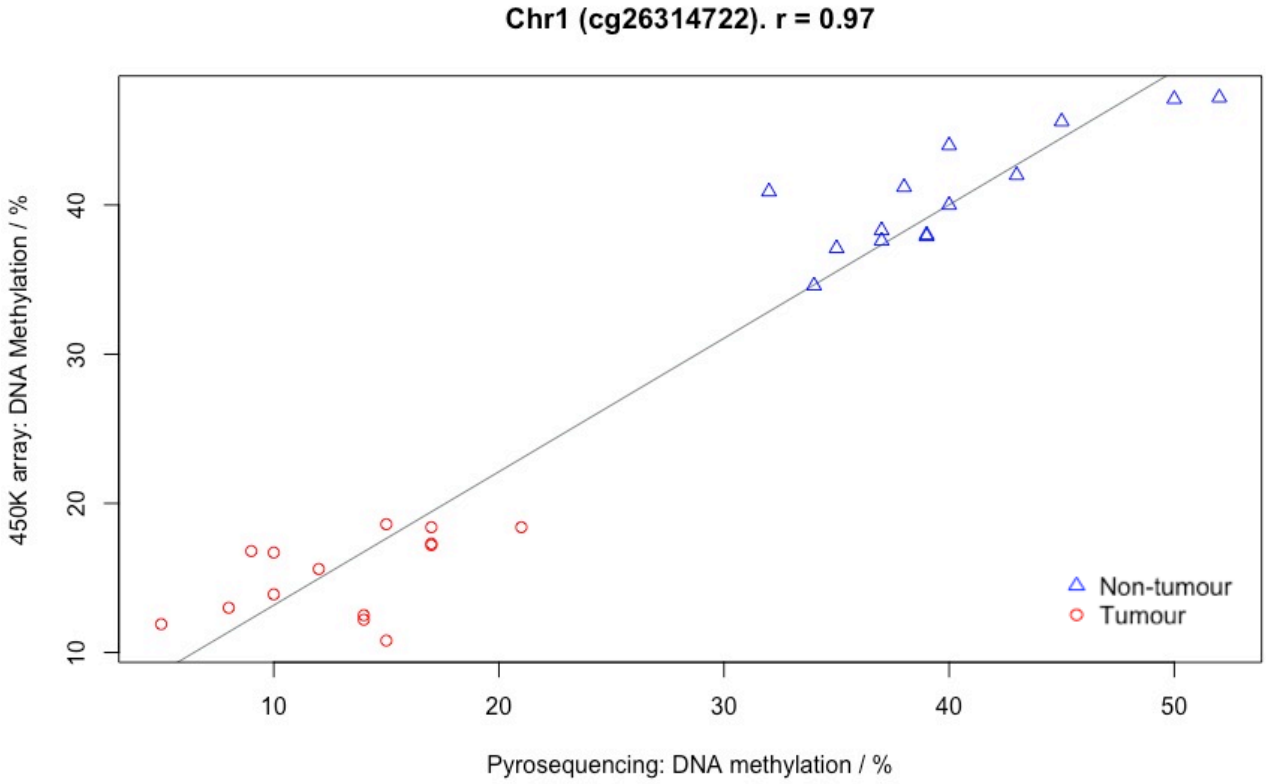

ZNF677 (cg24685755). r = 0.90
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour
20
40
60
80

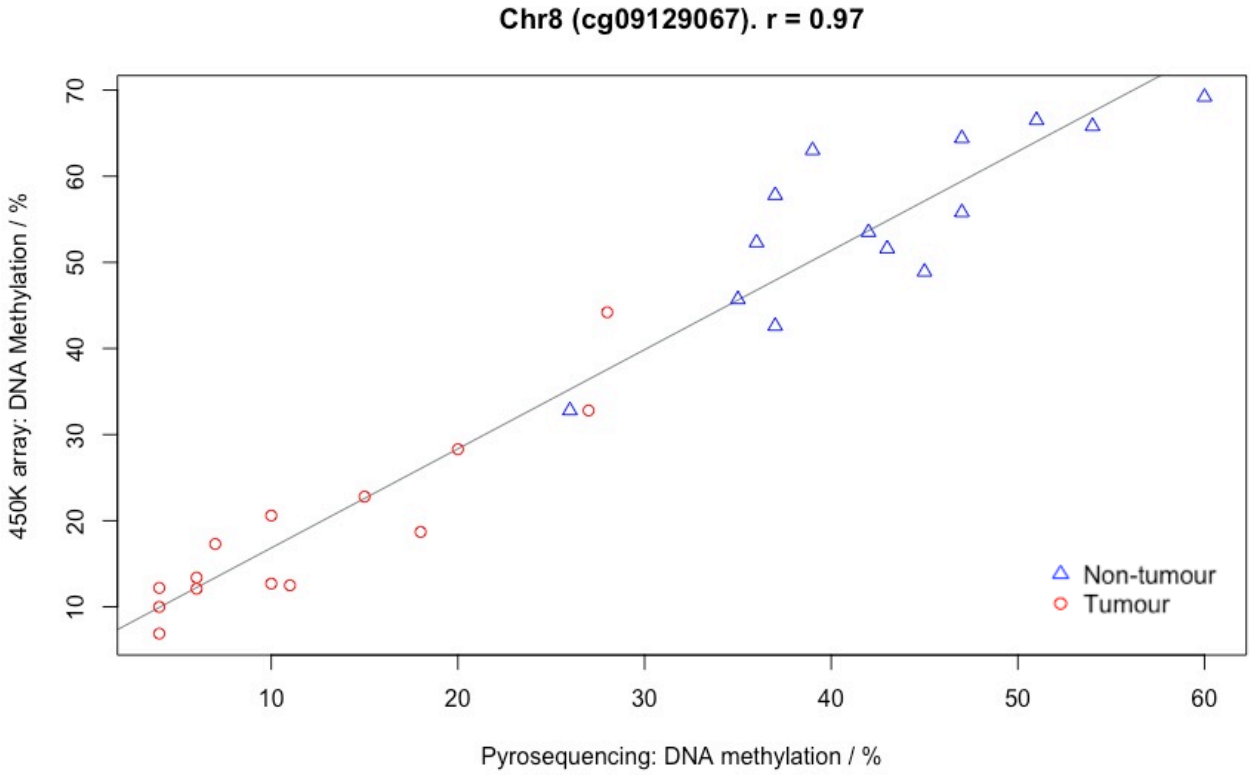

CCDC8 (cg03576469). r = 0.97
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour


LOC283999 (cg26034516). r = 0.87
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour


SPOCK_1 (5, cg24847829). r = 0.91
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour
20
40
60
80


SPOCK_1 (7, cg14650610). r = 0.87
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour
0
20
40
60
80


Chr1 (cg26314722). r = 0.97
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour
10
20
30
40
50


Chr8 (cg09129067). r = 0.97
450K array: DNA Methylation / %
Pyrosequencing: DNA methylation / %
Non-tumour
Tumour
10
20
30
40
50
60
70

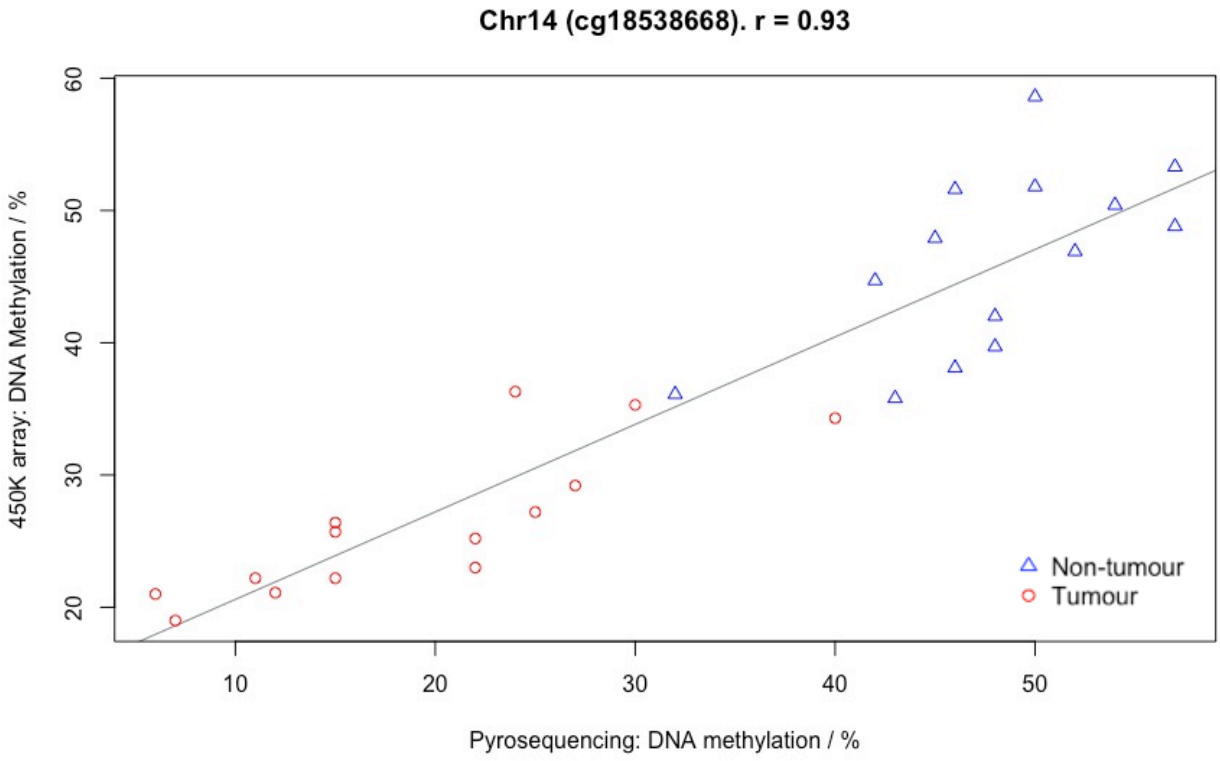


Figure 6.6: Scatterplots to demonstrate the correlation between methylation values of DMPs assessed by pyrosequencing and the Illumina 450k genome wide assay following the exclusion of samples obtained from patient R03. r, Pearson's correlation coefficient.

6.4 Methylation differences between tumour and normal mucosa

6.4.1 Differential methylation

Significant differences were observed between the methylation values of matched tumour and adjacent mucosa for all DMPs assessed (Figure 6.7). These results parallel those obtained from the Illumina 450k array.

Leong et al. classified hypermethylated tumour samples as those with DNA methylation values greater than two standard deviations (2σ) above the mean (μ) of non-tumour samples. Hypomethylated tumour samples were defined as those with DNA methylation values less than two standard deviations below the mean of non-tumour samples. Most samples analysed in our cohort met these criteria (Table 6.2). *CCDC8* DNA methylation demonstrated the lowest predictive value, with only four of nine samples reaching the required methylation level characteristic of tumour tissue. T001, T008, T010 and T016 each failed to meet these definitions of DNA hyper/hypomethylation in more than one assay. N013 and N001 displayed methylation levels characteristic of tumour samples in the *SPOCK*_1 (CpG 5) and Chr14 assays respectively.

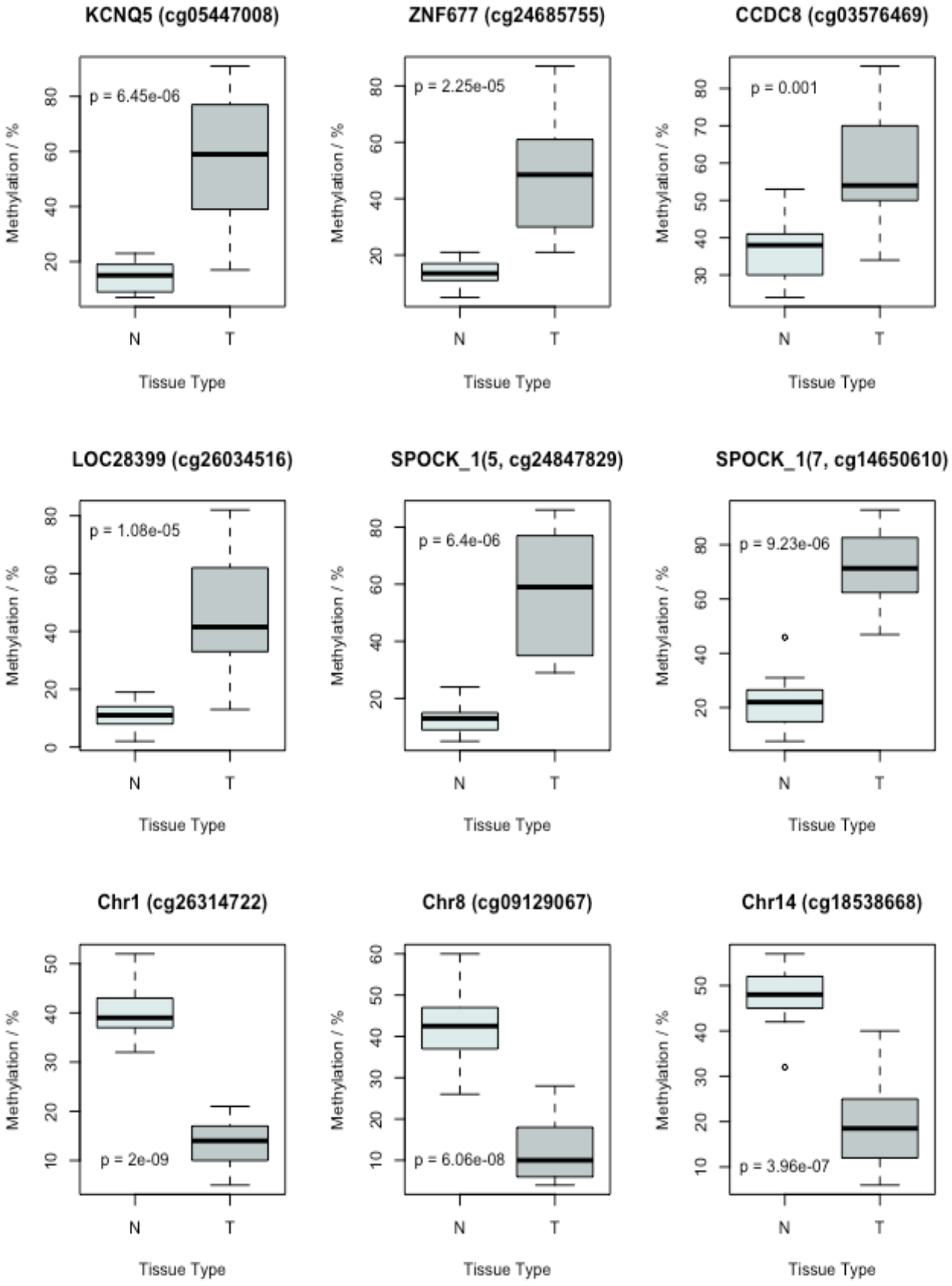


Figure 6.7: Boxplots to demonstrate the median difference in methylation between tumour (T) and non-tumour/adjacent mucosa (N) tissue samples. Whiskers represent 1.5 time the interquartile range or min/max values.

	T samples ≥ μ + 2σ of N / total	T samples < μ + 2σ of N	N samples ≥ μ + 2σ of N / total	N samples ≥ μ + 2σ of N
KCNQ5	13/14	T008	0/14	-
ZNF677	12/14	T001, T016	0/14	-
LOC283999	13/14	T010	0/14	-
CCDC8	4/9	T001, T004, T008, T010, T011	0/9	-
SPOCK_1(5)	13/13	-	1/13	N013
SPOCK_1(7)	13/13	-	0/13	-
	T samples ≤ μ - 2σ of N / total	**T samples > μ - 2σ of N**	**N samples ≤ μ - 2σ of N / total**	**N samples ≤ μ - 2σ of N**
Chr1	14/14	-	0/14	-
Chr8	12/14	T007, T016	0/14	-
Chr14	13/14	T016	1/14	N001

Table 6.2: Number of samples containing DNA methylation values greater/less than the mean of adjacent mucosa. T, tumour; N, non-tumour/adjacent mucosa; μ, mean; σ, standard deviation

6.4.2 Methylation of surrounding CpGs

Individual DMPs selected for validation were identified using the Illumina 450k assay. In order to determine whether the DNA methylation changes observed at target DMPs selected for this study are representative of changes in nearby CpG sites, the coefficient of variation (σ/μ) between individual CpGs was calculated for each pyrosequencing assay (Table 6.3) and illustrated in Figure 6.8. Due the low density of CpGs within the Chr1 assay, DNA methylation data was available for the target DMP only.

As illustrated in Figure 6.8, the degree of variation observed between individual CpGs was small, particularly with regard to tumour samples. The pattern of DNA hyper/hypomethylation between tumour tissue and adjacent mucosa was maintained across all CpG sites, indicating that the pattern is likely representative of the surrounding genome. LOC283999 and *SPOCK*_1 assays displayed the greatest degree of variation in non-maligant tissue, whereas Chr14 displayed the greatest variation in tumour tissue.

	σ of all CpGs: Non-Tumour	Coefficient of variance: Non-Tumour	σ of all CpGs: Tumour	Coefficient of variance: Tumour
KCNQ5	5.5	0.29	7.1	0.12
ZNF677	2.5	0.15	4.8	0.09
LOC283999	9.1	0.56	4.9	0.10
CCDC8	7.2	0.16	5.2	0.08
SPOCK_1	3.9	0.32	5.3	0.10
Chr8	2.8	0.06	1.0	0.07
Chr14	10.8	0.19	11.0	0.41

Table 6.3: Standard deviation (σ) and coefficient of variance (σ/μ) of the average methylation of individual CpGs within each pyrosequencing assay.

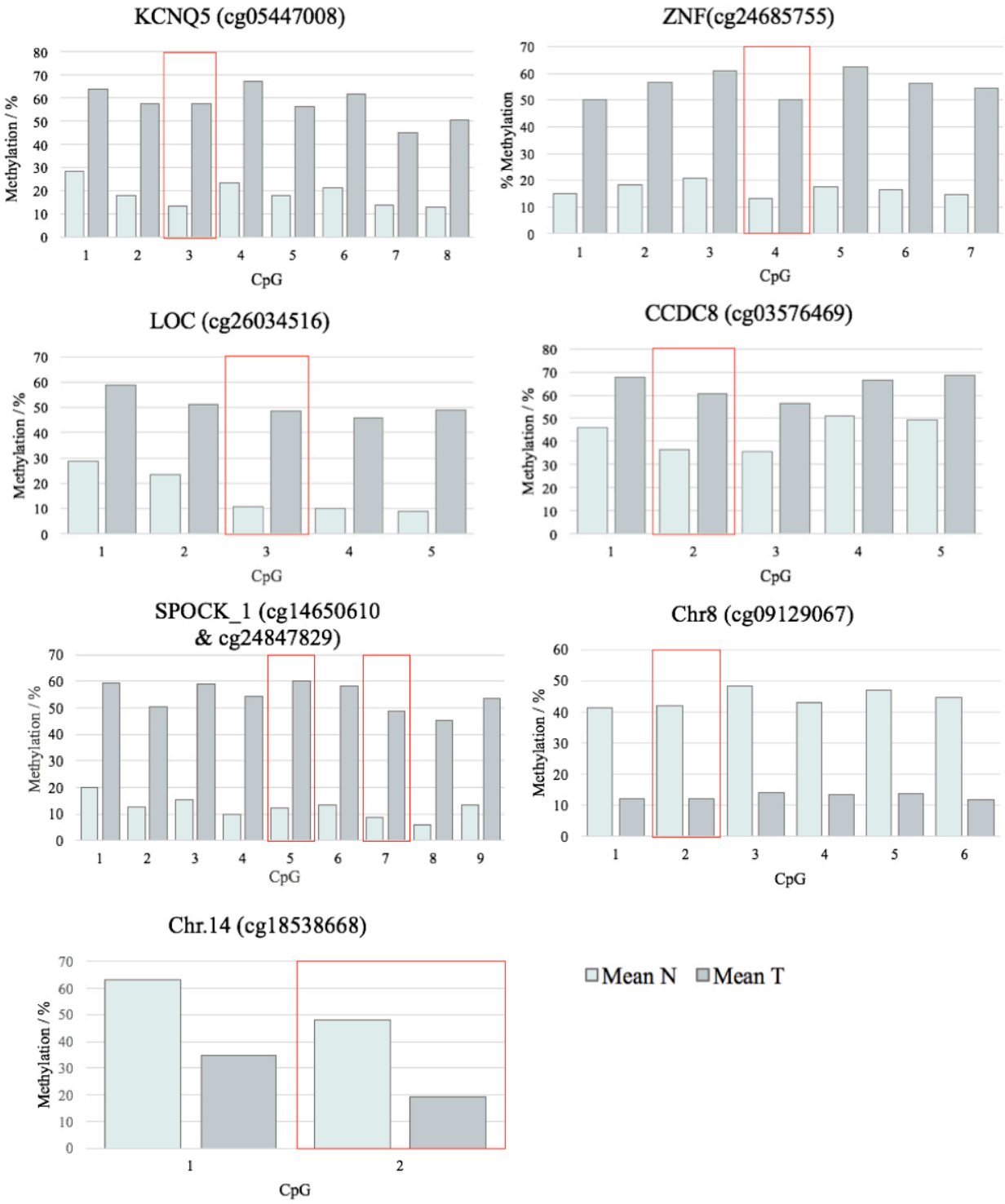


Figure 6.8. Plots to show the average methylation of all CpGs assessed by each pyrosequencing assay. Red boxes highlight the target DMP/s of each assay. T, tumour; N, non-tumour/ adjacent mucosa.

7 Results: Global DNA Methylation

7.1 Global DNA methylation analysis of RIST samples

Application of the LUMA assay to the ten eligible RIST samples demonstrated significantly lower global DNA methylation in tumour (mean 54.6%) than in adjacent tissue samples (mean 63.5%, P = 0.021, Figure 7.1).

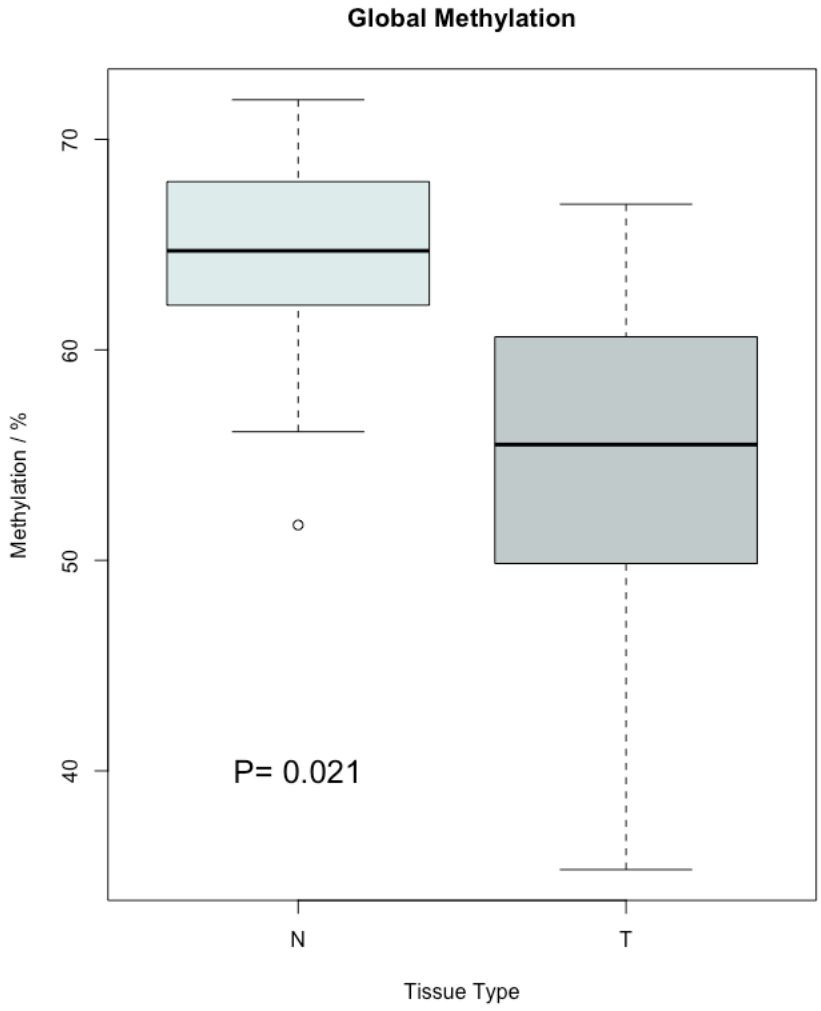


Figure 7.1: Median global DNA methylation levels of tumour (T) and non-tumour/ adjacent mucosa (N) tissue samples. Whiskers represent 1.5 time the interquartile range or min/max values. A paired t-test was used to compare global DNA methylation levels between tumour and adjacent mucosal.

8 Results: Application of Assays to Validation Set

8.1 Introduction

Preliminary studies and subsequent validation was performed in a relatively small number of tissue sample pairs (n=15). Pyrosequencing assays confirmed that the DNA methylation data obtained by 450k analysis was reproducible in the same sample cohort and are therefore likely to represent true biological variation. In order to determine whether these findings could be reproduced in an independent sample cohort, replication was performed using historically acquired sample pairs. Increased sample size and the availability of clinical outcome data in this cohort would also enable deeper exploration of the relationship between DNA methylation patterns and clinicopathological features.

8.2 Selection of tissue samples

8.2.1 Historically acquired samples

Matched tumour and adjacent mucosa tissue samples collected from 87 patients were obtained from the Exeter Tissue Bank (Section 4.2.1). Sufficient tissue for histological assessment and DNA extraction was available from 82 patients.

8.2.3 Histological evaluation of tumour load

Some studies have used predetermined minimum tumour percentage for inclusion in subsequent analysis. However, few have given any tangible justification. As histological assessment of RIST samples had been prohibited by small tissue volumes, we aimed to investigate the true value of histological assessment in the determination of DNA methylation level. We hypothesised that exclusion on the basis of tumour load[5] of adjacent tissue may be a flawed method as a result of tumour heterogeneity. To test this hypothesis, we determined the histological tumour load of immediately adjacent tissue to that

[5] Here we define tumour load as the percentage of a histological specimen occupied by tumour cells, to the nearest 5%, as determined by an experienced consultant histopathologist.

analysed and sought to assess the impact of this on DNA methylation values observed.

82 samples were dissected and histologically assessed as described in Section 4.3. 13 samples contained no tumour cells and/or were necrotic. One adjacent tissue sample contained all soft tissue with no epithelium. Therefore, histologically confirmed tumour was present in 68 sample pairs (total n=136). Remaining samples contained a mean tumour load of 67.8% (±29.2%, Figure 8.1). Tumour load greater than or equal to 80% was present in 39 samples. Tumour load less than or equal to 40% was present in 19 samples.

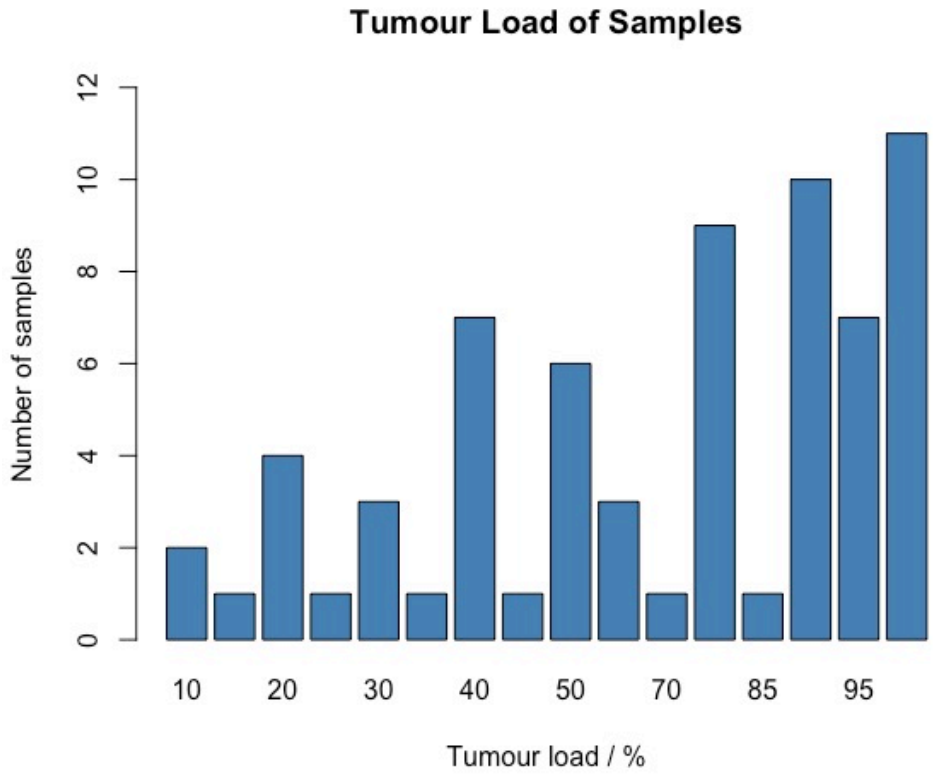


Figure 8.1: Histogram to illustrate the histologically confirmed tumour load of historically acquired tissue samples. We define tumour load as the percentage of a histological specimen occupied by tumour cells, to the nearest 5%, as determined by an experienced consultant histopathologist.

8.3 Pyrosequencing

8.3.1 Correlation between DNA methylation values of RIST and historically acquired samples

Whereas the RIST cohort comprised entirely of patients with rectal cancer, historically acquired samples were obtained from patient with tumours of both the colon and rectum. We therefore aimed to assess whether the DNA methylation levels determined in these cohorts were representative of one another. The correlation between average DNA methylation in each cohort was calculated using Pearson's test. As demonstrated in Figure 8.2, a strong linear relationship between the two cohorts was observed (r=0.98-0.99).

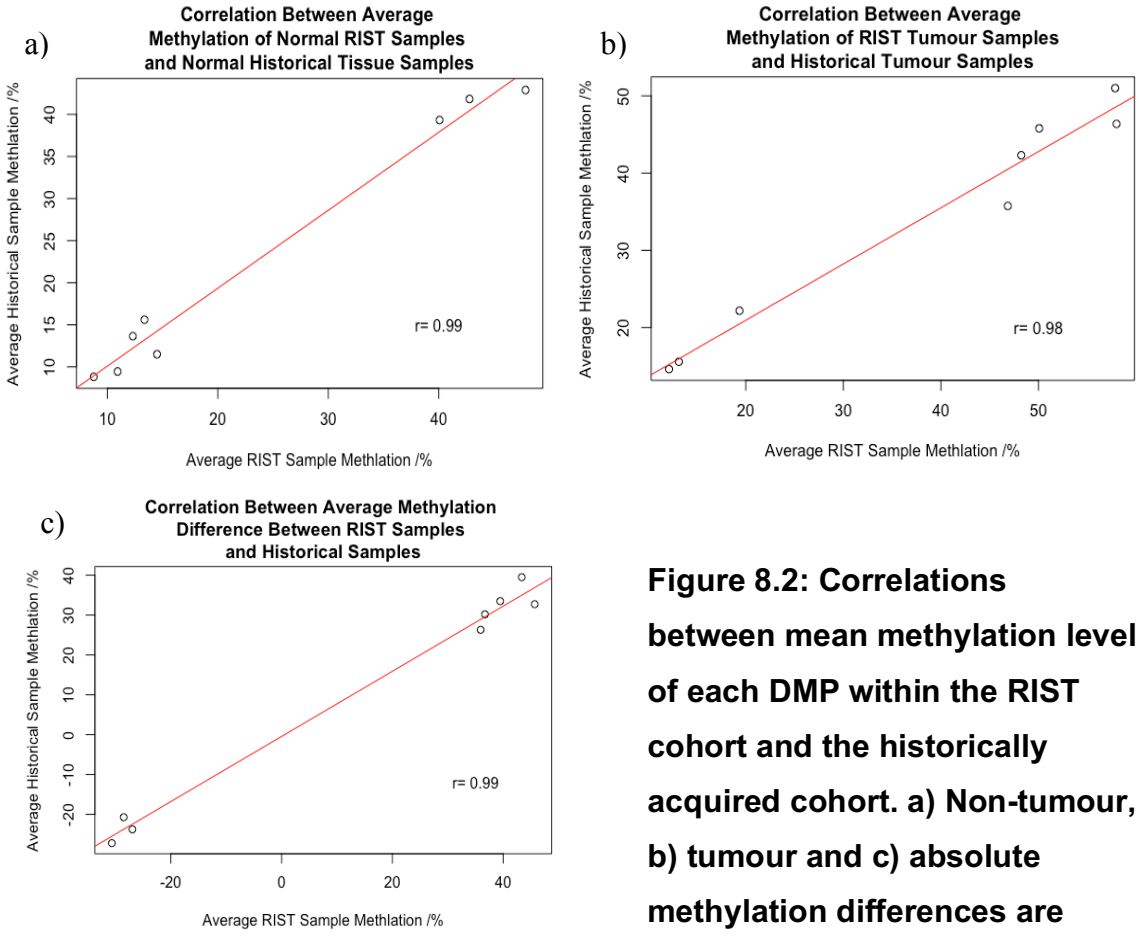


Figure 8.2: Correlations between mean methylation level of each DMP within the RIST cohort and the historically acquired cohort. a) Non-tumour, b) tumour and c) absolute methylation differences are plotted separately. Each point represents a separate assay. r, Pearson's correlation coefficient.

Little difference was observed between the mean DNA methylation values of non-tumour tissues of each cohort. In contrast, mean DNA methylation was slightly higher in tumour samples of the RIST cohort than those of the historical cohort. The greatest variation was observed in LOC283999 (RIST 46.9% vs. Historical 35.7%, P = 0.075, 95% CI= -23.5 - 1.3) and the fifth CpG of *SPOCK*_1 (RIST 58.0% vs. Historical 46.4%, P = 0.090, 95% CI= -25.3-2.0). These discrepancies may result from the inclusion of colon cancers within the historical acquired cohort.

8.3.2 Association between tumour load and DNA methylation

In order to determine the relationship between tumour load and DNA methylation levels, correlations were calculated using Pearson's test. We hypothesised that absolute tumour methylation, as well as methylation difference between tumour and adjacent tissue, would be greater in tissue samples with high tumour load. Strong correlations may then be used to justify the use of a lower tumour percentage threshold for inclusion in the study.

The relationship between tumour load and DNA methylation was stronger at some loci, such as *SPOCK*_1and *ZNF677* than others (Figure 8.3). However high and low levels of DNA methylation were present across all levels of tumour cellularity. Therefore, we did not feel that any of the assays displayed sufficient predictive power to exclude any samples from inclusion in the study. Similar relationships were also observed between tumour load and DNA methylation difference between tumour and adjacent tissue (Figure 8.4).

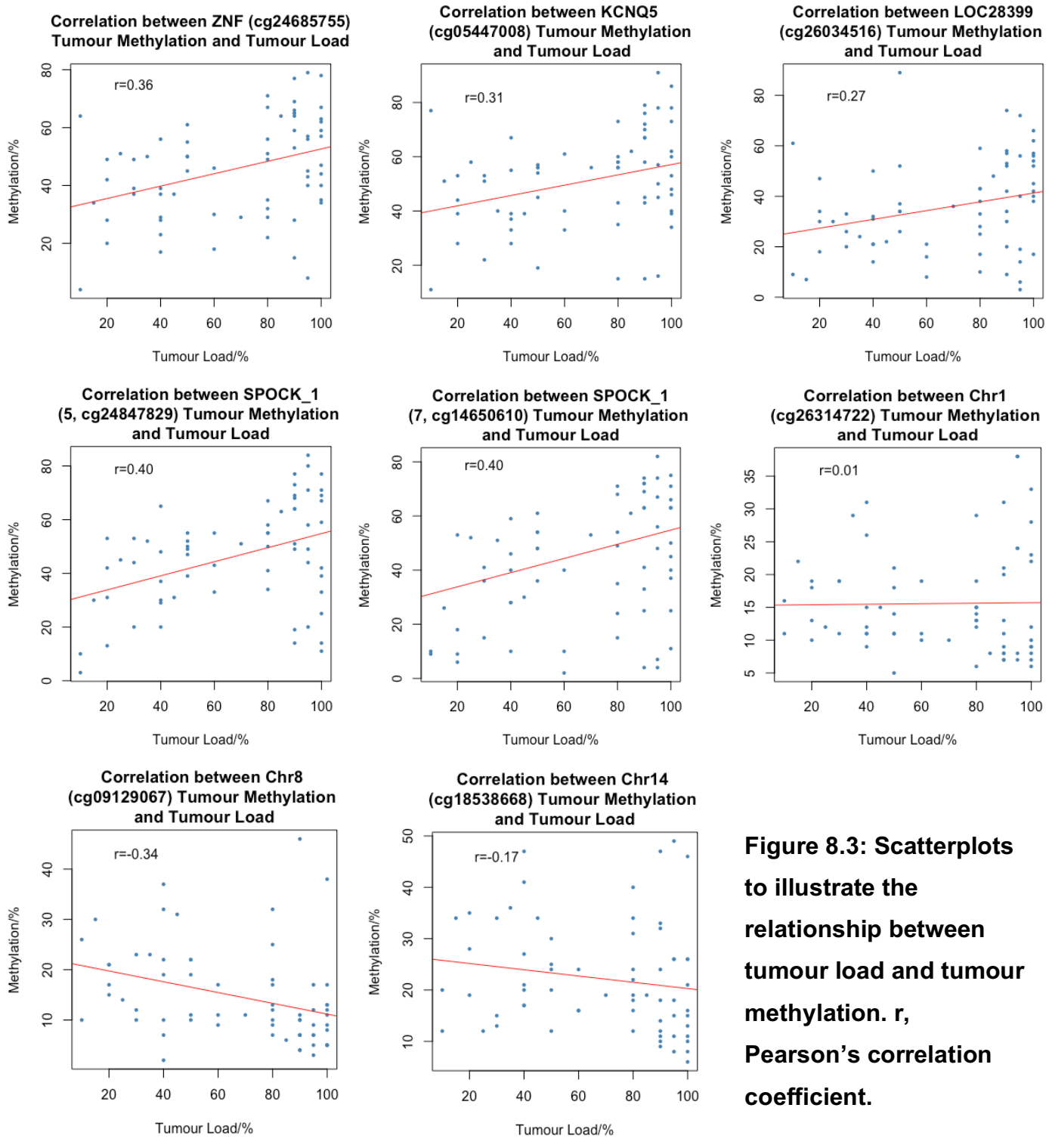


Figure 8.3: Scatterplots to illustrate the relationship between tumour load and tumour methylation. r, Pearson's correlation coefficient.

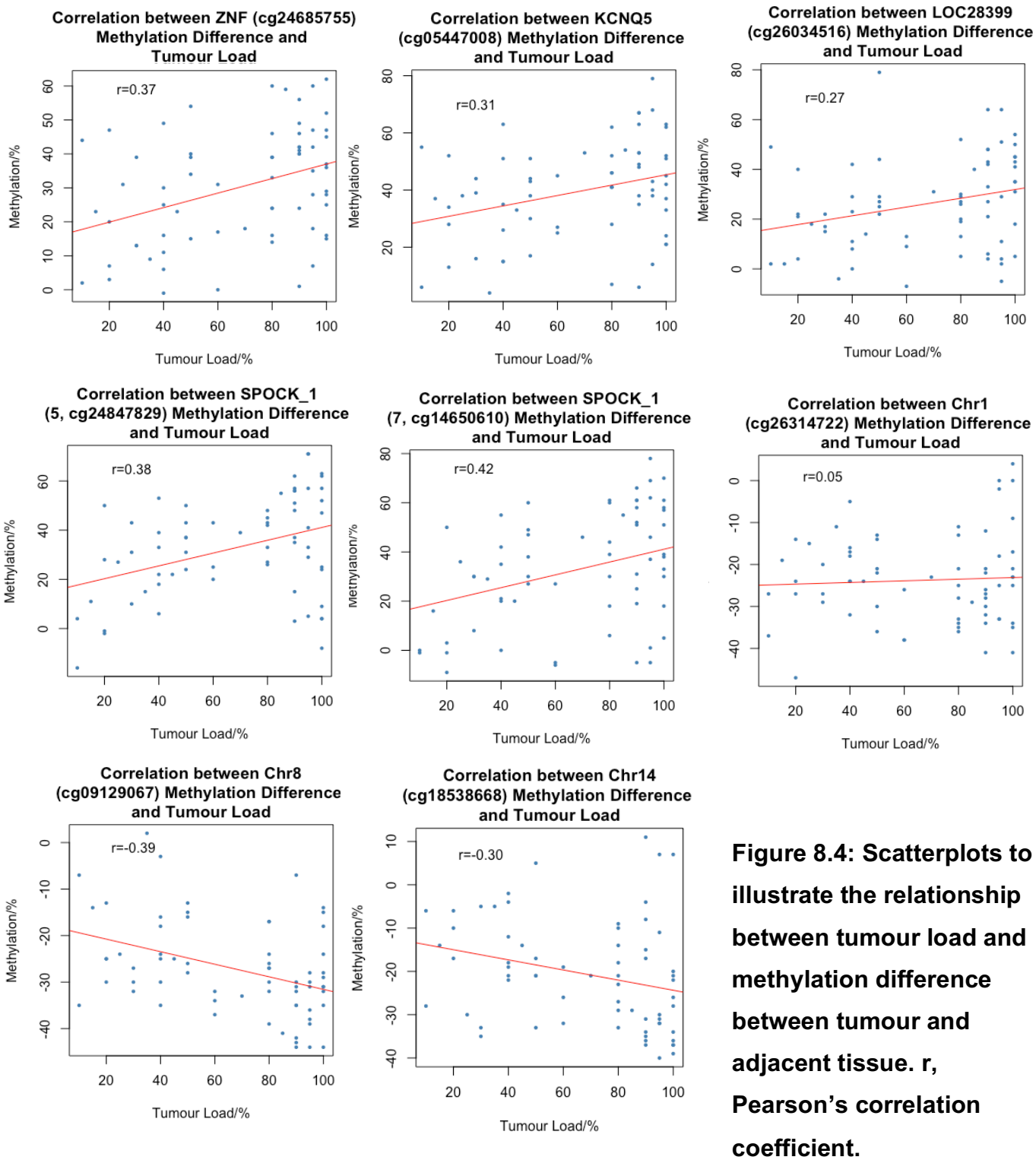
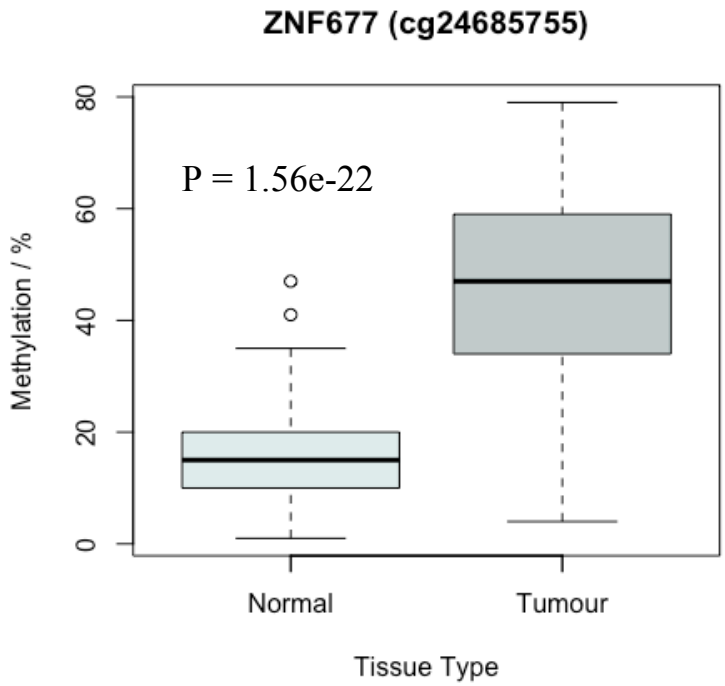
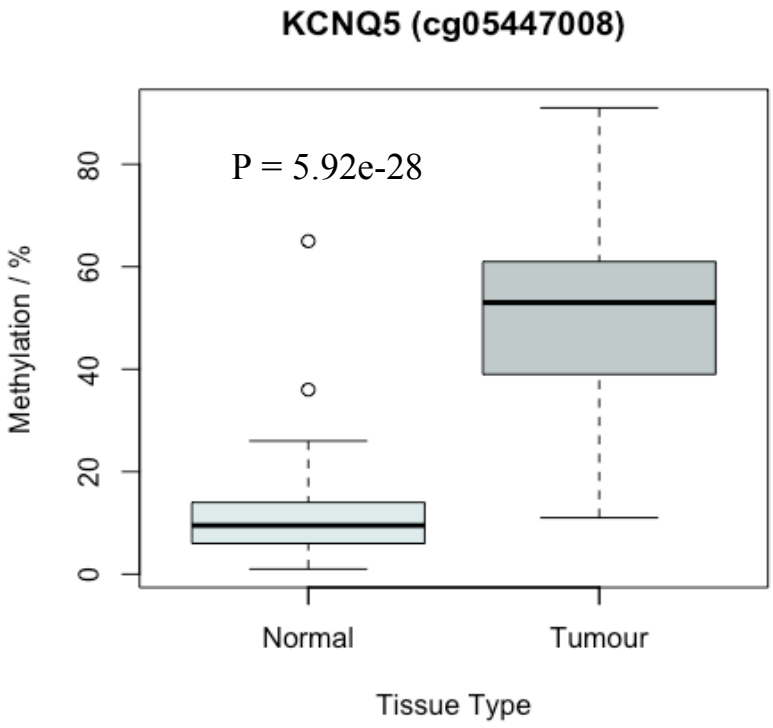


Figure 8.4: Scatterplots to illustrate the relationship between tumour load and methylation difference between tumour and adjacent tissue. r, Pearson's correlation coefficient.

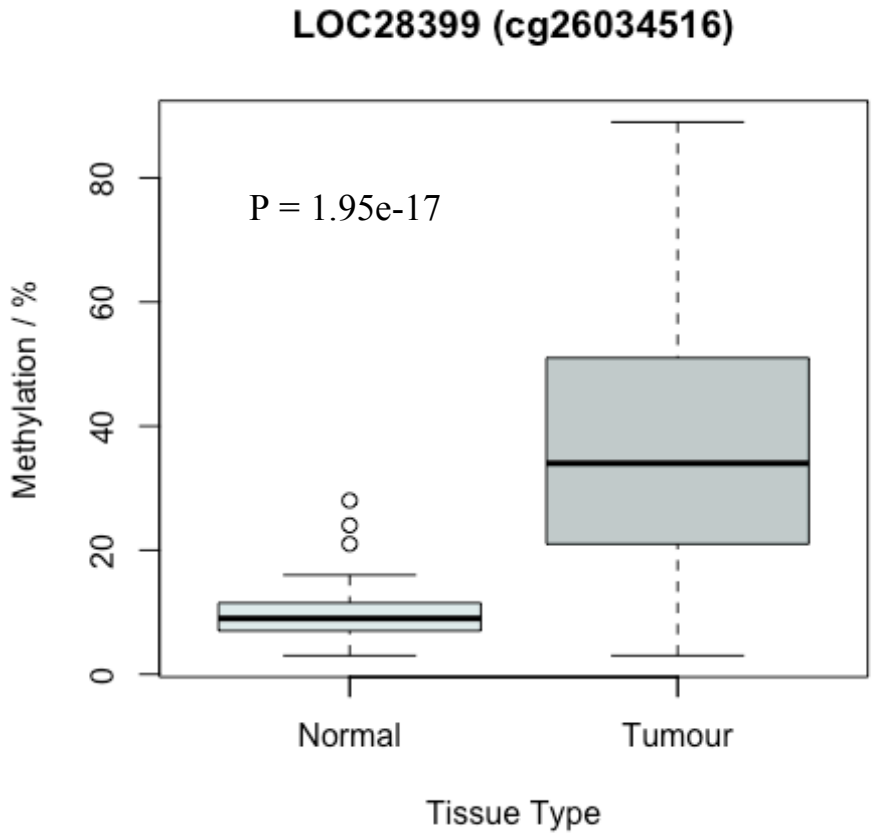
8.3.3 Differences between tumour and adjacent tissue

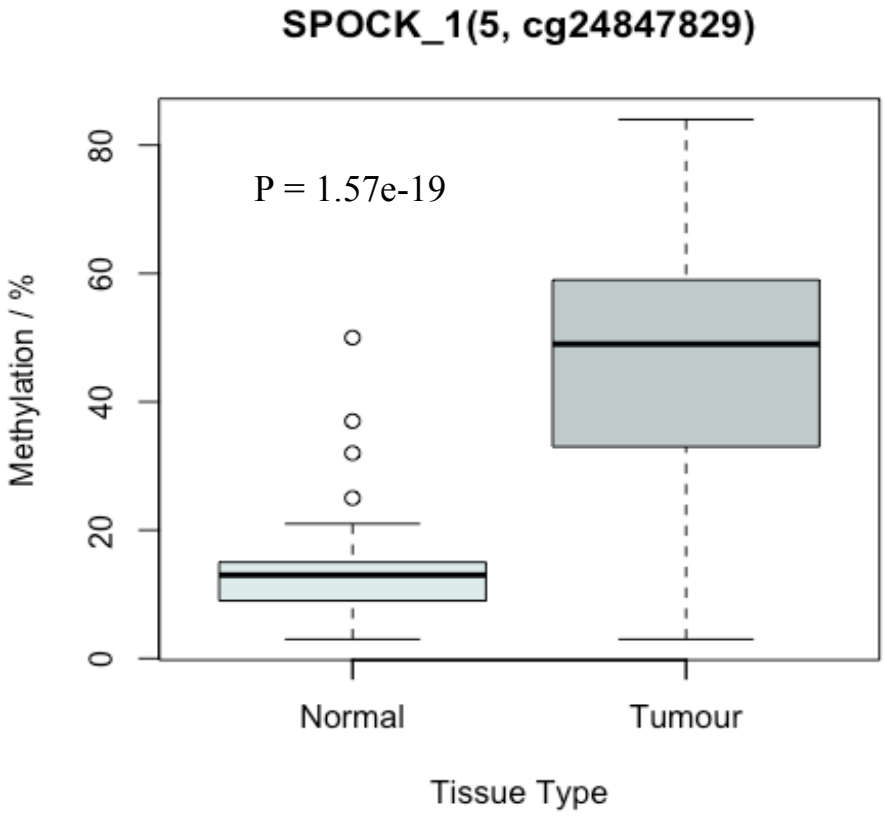
Aim to determine whether these differences remained in the historically acquired cohort. DNA methylation levels of each target CpG in tumour and adjacent tissue were compared using paired t-tests (Figure 8.5)

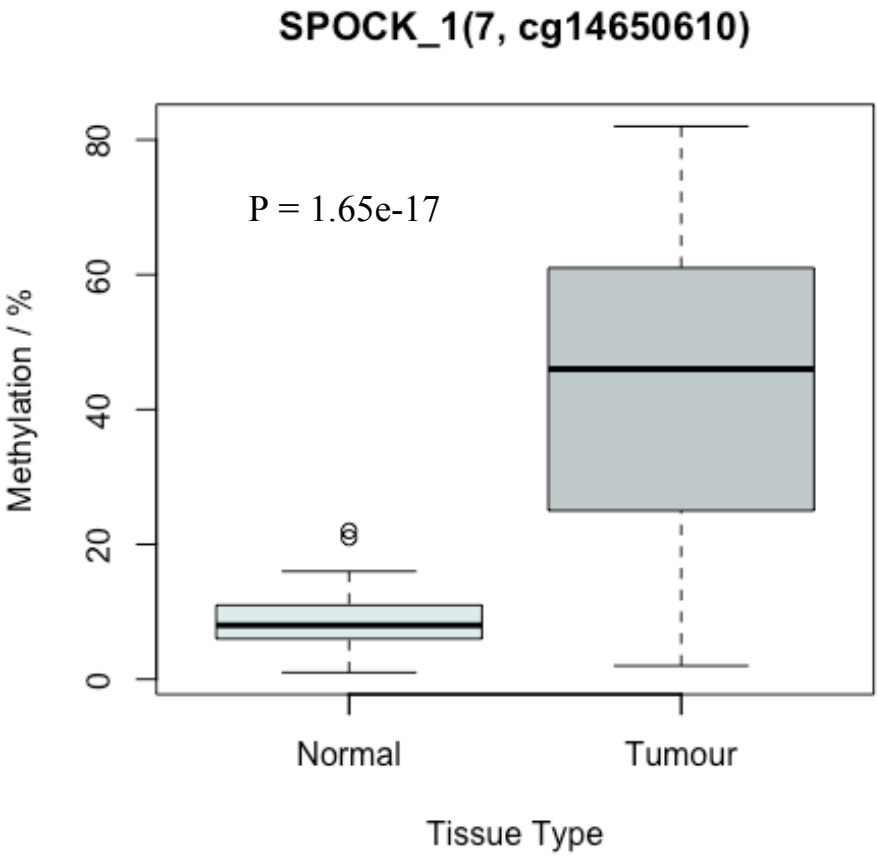


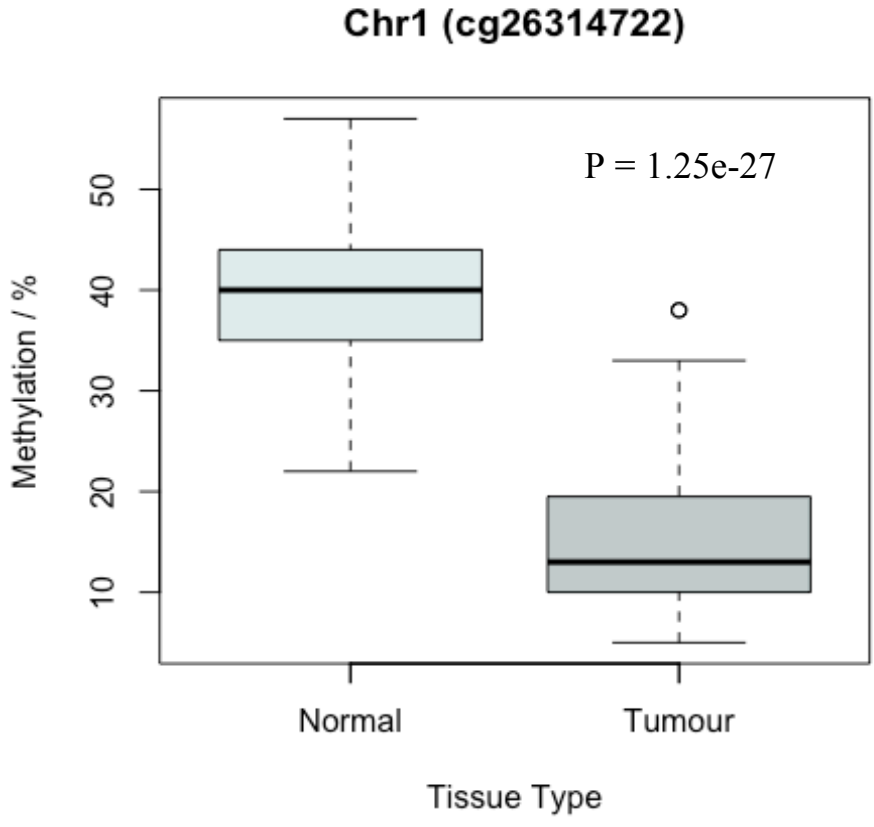


.


LOC28399 (cg26034516)
P = 1.95e-17
Methylation / %
0
20
40
60
80
Normal
Tumour
Tissue Type


SPOCK_1(5, cg24847829)
P = 1.57e-19
Methylation / %
0
20
40
60
80
Normal
Tumour
Tissue Type


SPOCK_1(7, cg14650610)
P = 1.65e-17
Methylation / %
0
20
40
60
80
Normal
Tumour
Tissue Type


Chr1 (cg26314722)
P = 1.25e-27
Methylation / %
10
20
30
40
50
Normal
Tumour
Tissue Type

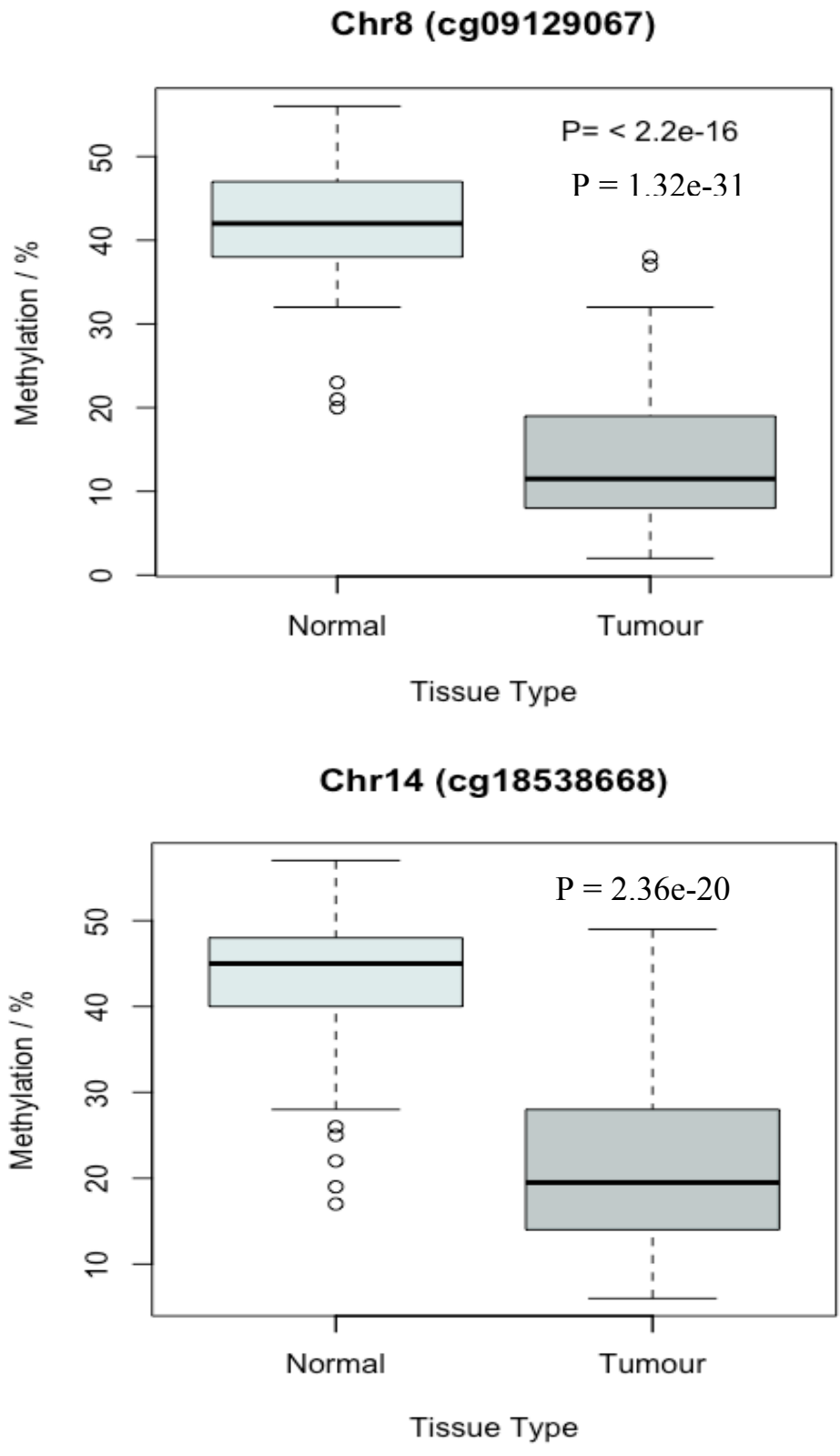


Figure 8.5: Boxplots to illustrate methylation differences between tumour tissue and adjacent mucosa. P values of paired t-tests are given. Boxes represent upper and lower quartiles and median DNA methylation values. Whiskers represent 1.5 times IQR or min/max values.

Raw DNA methylation data obtained from pyrosequencing is given in Appendix 3. All CpGs of interest displayed highly significant levels of differential methylation between tumour and adjacent mucosal tissue (Table 8.1). The range of DNA methylation values present in tumour tissues was greater than that of adjacent tissue which may represent differences in tumour characteristics.

	ZNF677	*KCNQ5*	LOC283999	*SPOCK(5)*	*SPOCK(7)*	Chr1	Chr8	Chr14
Samples anaylsed	67	66	67	65	65	67	66	66
t	14.7	18.8	11.5	12.9	11.7	18.4	21.8	-13.4
P-value	1.56e-22	5.92e-28	1.95e-17	1.57e-19	1.65e-17	1.25e-27	1.32e-31	2.36e-20
95% CI	26.1-34.3	35.3-43.7	21.7-30.9	27.7-37.8	27.8-39.2	-26.3--21.2	-29.7--24.7	-23.8--17.6
Average DNA methylation Difference	30.2	39.5	26.3	32.7	33.5	-23.7	-27.2	-20.7

Table 8.1: Number of successfully analysed samples and summary statistics of paired t-tests for each DMP.

DNA methylation levels of individual samples are visualised in Figure 8.6. Hierarchical clustering successfully discriminated tumour sample assays from adjacent tissue samples. No clear association was observed with relation to Dukes' stage or tumour site in this plot. Absolute change in DNA methylation level between tumour and adjacent tissue varied considerably between patients. In addition, some variation was also observed in the pattern of hyper/ hypomethylated CpGs within the same sample. This variation may represent differences in clinicopathological features of each sample or may result from tumour heterogeneity.

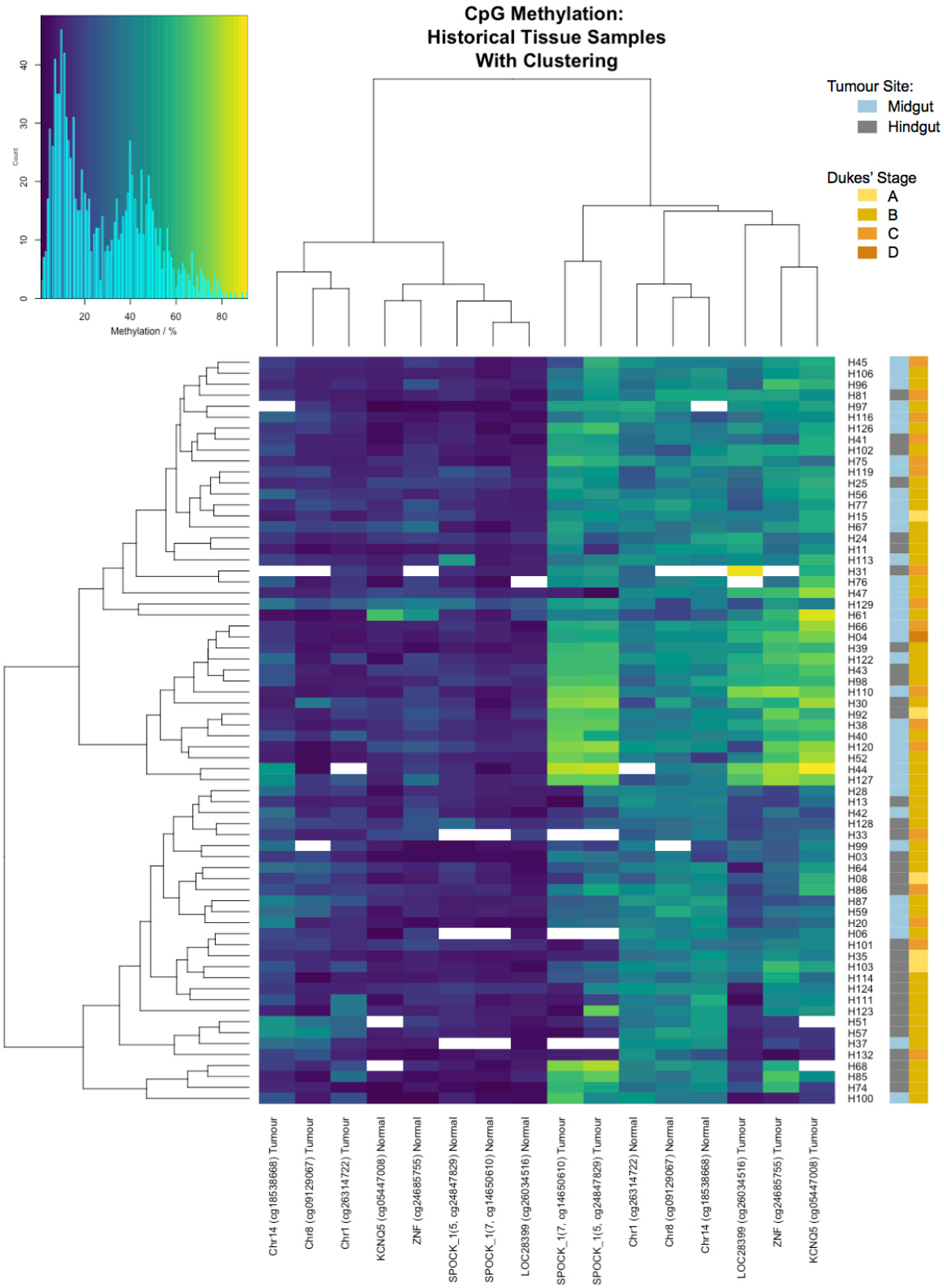


Figure 8.6: Heatmap to demonstrate methylation values of individual tissue samples at each CpG site. Hierarchical clustering successfully discriminated tumour sample assays from adjacent tissue samples. No clear association was observed with relation to Dukes' stage or tumour site in this plot.

8.3.4 Predictive power of selected DMPs as a panel

Interpatient and intratumour heterogeneity between individual DMP methylation levels, may limit their ability to distinguishing tumour from adjacent tissue when used in isolation. We therefore aimed to create a model with the ability to predict tumour status using combined DNA methylation data from a panel of CpGs.

DNA methylation data from the historical tissue cohort was used as a training set to create a logistic regression model for the prediction of tissue type (tumour or adjacent mucosa). Due to strong collinearity between regressors, no single DMP could significantly describe the model. Five of eight DMPs significantly added to the model when added sequentially (Table 8.2). Data from all DMPs were included in the final model. Significantly increased ability to distinguish tumour from adjacent tissue was observed in comparison to a null model defined by the intercept only (P = 8.03E-27).

	Deviance	Residual Deviance	Pr (>Chi)
ZNF677	87.95	72.86	< 2.20E-16
KCNQ5	20.66	52.20	5.48E-06
SPOCK_1 (5)	8.05	44.15	4.55E-03
SPOCK_1 (7)	2.92	41.23	0.09
LOC283999	0.53	40.70	0.46
Chr.1	16.31	24.39	5.39E-05
Chr.8	5.35	19.04	0.02
Chr.14	0.49	18.55	0.48

Table 8.2. Contribution of DMPs to the model when sequentially added.

The model was applied to RIST samples for which complete DNA methylation data was available (n=26). The model successfully predicted tumour status in 100% of cases with a high degree of confidence (Figure 8.7)

.

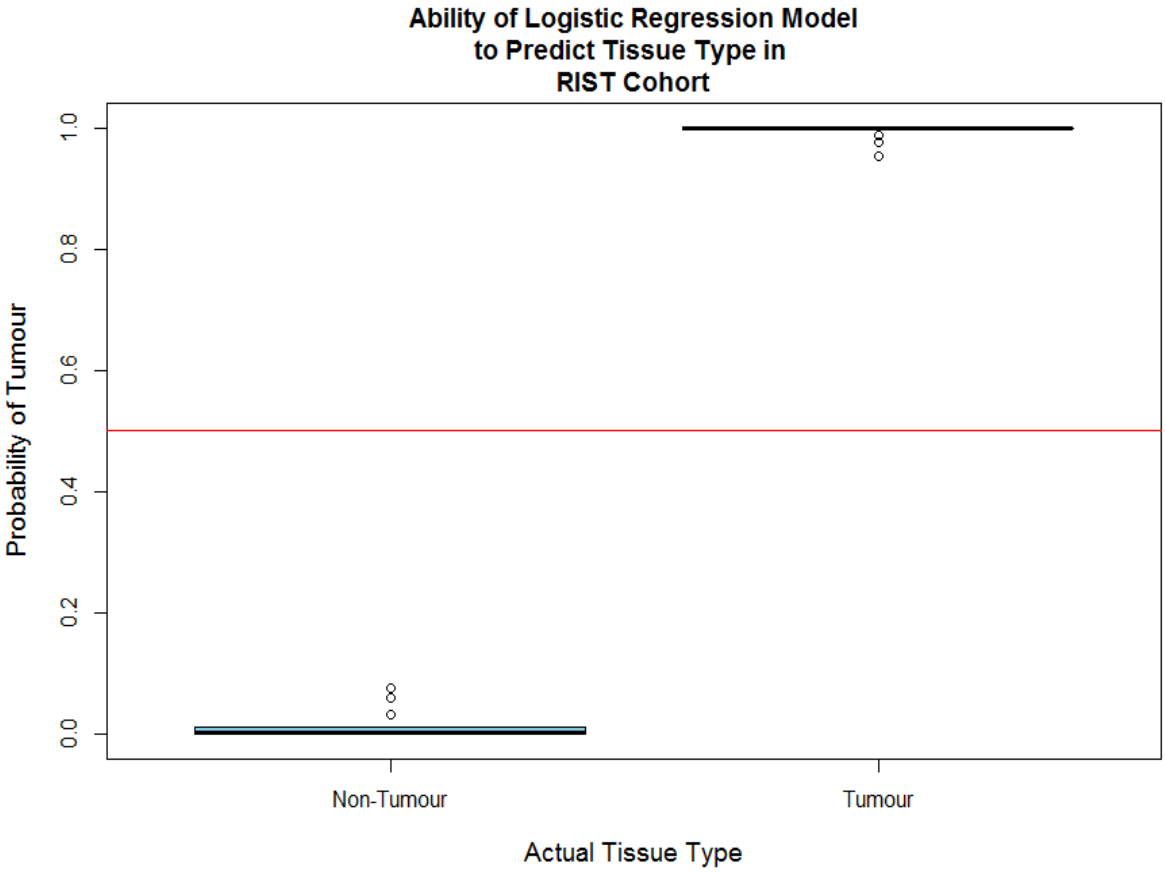


Figure 8.7: The ability of a logistic regression model to predict tissue type in the RIST cohort. Probabilities greater than 0.5 (red line) correspond to likely tumour samples. In this case the model was able to successfully predict tumour vs. non- tumour in 100% of cases.

9 Results: Associations between clinicopathological data and DNA methylation

9.1 Univariate analyses

9.1.1 Introduction

A greater understanding of the relationships between DNA methylation and clinicopathological features may provide insight into the biological impact of DMP methylation. Initial investigations were performed through univariate analysis of each DMP and cliniopathological feature (Table 4.2) in turn.

Only one and seven sample pairs were obtained from patients with tumours of T-stage one and two respectively. These groups were therefore combined for analysis. Similarly, only four patients in the cohort possessed N2 tumours. N stage was therefore considered a binary variable based on the presence or absence of nodal involvement. Right and left sided tumours were defined as those proximal and distal to the splenic flexure respectively. Tumours of the descending (n=5) and sigmoid colon (n=21) were grouped together with tumours of the rectum (n=9) under this definition. Most tumours were of Dukes' stage B (n=45) or C (n=17). Dukes' stages A and B, and stages C and D were therefore combined to form two groups. Insufficient group sizes were available for meaningful analysis of tumour metastasis or histological tumour type.

Univariate analysis was conducted using unpaired t-tests and ANOVA for binary and polytomous variables respectively. Each analysis was conducted for tumour tissue, adjacent mucosa and methylation difference in turn.

9.1.2 Associations between selected CpGs and tumour characteristics

Table 9.1 summarises P-values for each of these analyses. Associations with P-values <0.05 are illustrated in Figure 9.1.

No associations were observed between DNA methylation and gender or DNA methylation and the presence of extramural vascular invasion (EMVI).

Tumour site was the most commonly significant feature—arising in seven of the tests performed. Midgut tumours displayed higher DNA methylation values than hindgut tumours at *ZNF677* and Chr1 in adjacent mucosa, and at *KCNQ5* and *SPOCK*_1(7) in tumour tissue. DNA hypomethylation of Chr1 in tumour tissue was associated with advanced Dukes' stage and lymph node involvement.

Of the 12 tests that reached significance in this study, three were in relation to DNA methylation difference between tumour and adjacent tissue. DNA methylation of the corresponding tumour samples also reached significance in all three of these cases. As methylation difference is dependent upon absolute tumour methylation, both tests are likely to reflect the same phenomenon.

These findings should be viewed with caution however, as in the 168 independent analyses performed 8.4 would be expected to reach statistical significance by chance alone (using 95% confidence intervals). The corrected P value to achieve statistical significance in this study, as determined by the Bonferroni method, was $2.98e^{-04}$. This was not met by any of the analyses performed however may be over-stringent given the probable correlation between variables assessed.

	ZNF677 N	*ZNF677* T	*ZNF677* D	*KCNQ 5* N	*KCNQ 5* T	*KCNQ 5* D	*SPOCK_ 1 (5)* N	*SPOCK_ 1 (5)* T	*SPOCK_ 1 (5)* D	*SPOCK_ 1 (7)* N	*SPOCK_ 1 (7)* T	*SPOCK_ 1 (7)* D	LOC 28399 N	LOC 28399 T	LOC 28399 D
Tumour site †	0.02	0.11	0.70	0.30	0.02	0.07	0.29	0.07	0.20	0.87	0.01	0.02	0.64	0.063	0.082
Gender †	0.49	0.40	0.61	0.82	0.89	0.79	0.52	0.92	0.89	0.15	0.69	0.89	0.97	0.16	0.17
EMVI †	0.39	0.97	0.78	0.35	0.62	0.44	0.37	0.88	0.70	0.19	0.59	0.47	0.69	0.67	0.63
Dukes' †	0.87	0.86	0.92	0.46	0.76	0.96	0.63	0.29	0.39	0.12	0.49	0.75	0.70	0.31	0.40
Differentiation †	0.03	0.16	0.71	0.48	0.22	0.54	0.38	0.86	0.78	0.94	0.61	0.65	0.08	0.17	0.22
Nodal involvement †	0.81	0.81	0.71	0.53	0.98	0.79	0.65	0.39	0.51	0.13	0.89	0.86	0.75	0.42	0.49
T Stage *	0.71	0.95	0.93	0.82	0.87	0.67	0.20	0.33	0.12	0.002	0.35	0.15	0.41	0.99	0.90
	Chr1 N	Chr1 T	Chr1 D	Chr8 N	Chr8 T	Chr8 D	Chr14 N	Chr14 T	Chr14 D						
Tumour site †	0.01	0.61	0.03	0.75	0.45	0.64	0.03	0.49	0.05						
Gender †	0.18	0.15	0.91	0.81	0.75	0.91	0.94	0.21	0.31						
EMVI †	0.48	0.72	0.79	0.13	0.45	0.87	0.24	0.72	0.38						
Dukes' †	0.51	0.01	0.03	0.97	0.46	0.65	0.54	0.31	0.72						
Differentiation †	0.62	0.63	0.60	0.26	0.68	0.24	0.30	0.72	0.57						
Nodal involvement †	0.34	0.01	0.01	0.94	0.36	0.56	0.52	0.24	0.64						
T Stage *	0.62	0.82	0.74	0.16	0.37	0.94	0.49	0.82	0.67						

Table 9.1: Summary of p-values obtained from univariate analysis of relationships between DNA methylation and cliniopathological features. P-values <0.05 are highlighted. EMVI, extramural vascular invasion; T, tumour; N, non-tumour/adjacent mucosa; D, methylation difference between tumour and adjacent tissue. t-test†, ANOVA*.

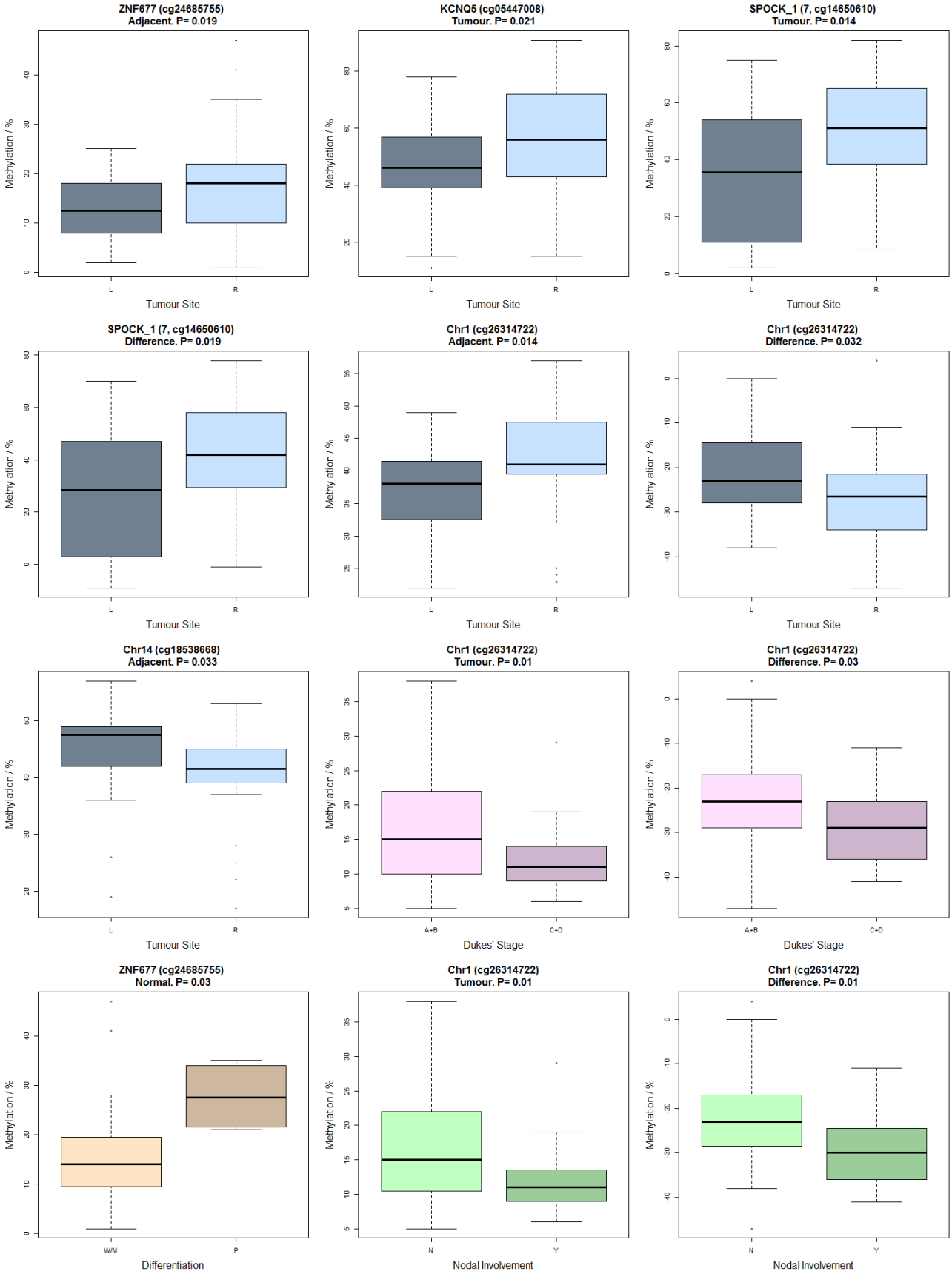


Figure 9.1: Boxplots of nominally significant relationships (P<0.05) between DMPs and clinicopathological features. Boxes are coloured according to variable analysed. L, left; R, right; W/M, well/moderately differentiated; P, poorly differentiated; N, no lymph node involvement; Y, positive lymph node involvement.

9.2 Multivariate analysis

Univariate analyses failed to provide any clear associations between DNA methylation and clinicopathological features. We therefore aimed to determine whether the incorporation of DNA methylation data from multiple CpG sites could be used to provide greater insight into these relationships.

Tumour site was the most commonly significant feature in univariate analyses and was therefore chosen for further investigation. As the RIST sample cohort comprised entirely of patients with rectal cancer, its use as a test set in the context of left vs right sided tumours would be of limited value. The historical cohort was therefore split into training and test sets containing 49 and 19 patients respectively. The training set was used to formulate two individual logistic regression models using DNA methylation data obtained tumour tissue (modT) and methylation difference (modD). All DMPs were incorporated into the model as predictors. Null models were defined by the intercept only in each instance.

ModT failed to outperform the null model (P = 0.21) and was therefore omitted from further analysis. ModD significantly outperformed the null model (P = 0.04). KCNQ5 was the only DMP significantly associated with tumour site (P = 0.03) and to contribute significantly to the model (P = 0.006).

Data from all DMPs were included in the final model and applied to the test set containing 16 complete cases (proximal tumours n=8, distal tumours n=8). Correct prediction of tumour site was possible in only 75% and 50% of left and right sided cases respectively (Figure 9.2).

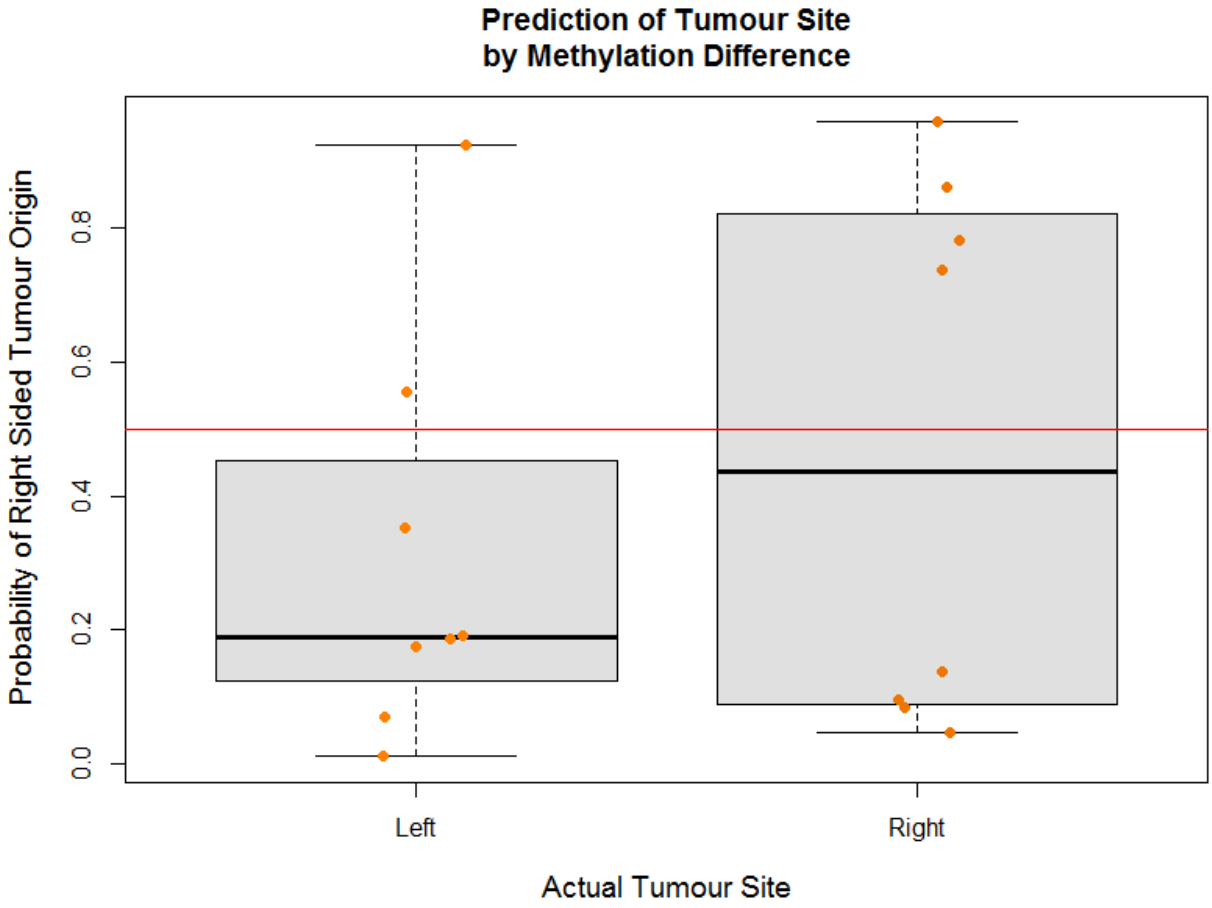


Figure 9.2: Ability of logistic regression model (modD) to predict tumour site. Probability values greater than 0.5 correspond to right sided tumours.

10 Results: Applicability of assays to FFPE samples

10.1 Introduction

FFPE is routinely performed in clinical practice to enable histological assessment of tissue. Through the formation of cross-links between the lysine residues of proteins, formalin fixation prevents the progression of enzymatic proteolysis within cells.(158) This enables the preservation of histological morphology and strength. However, FFPE also results in DNA fragmentation and the formation of sequence artefacts which renders genetic analysis of these samples difficult.(159)

DNA obtained from FFPE samples in our preliminary studies failed to reach sufficient quality for assessment by Illumina 450k analysis. We therefore aimed to determine the applicability of our pyrosequencing assays to genomic DNA obtained from these samples.

10.2 PCR of DNA from FFPE samples

DNA samples were obtained from 22 formalin fixed paraffin embedded paired tissue samples. Tissue processing and DNA extraction had been performed in advance.

Samples were diluted and bisulphite converted as described in Section 4.6. Three FFPE DNA samples were randomly selected and amplified using PCR primers of each assay. 100% methylated HeLA DNA was used as a control in each case. Identification of successful assays was determined by the presence of appropriately sized amplicons by gel electrophoresis as described in Section 4.8.2 (Figure 10.1).

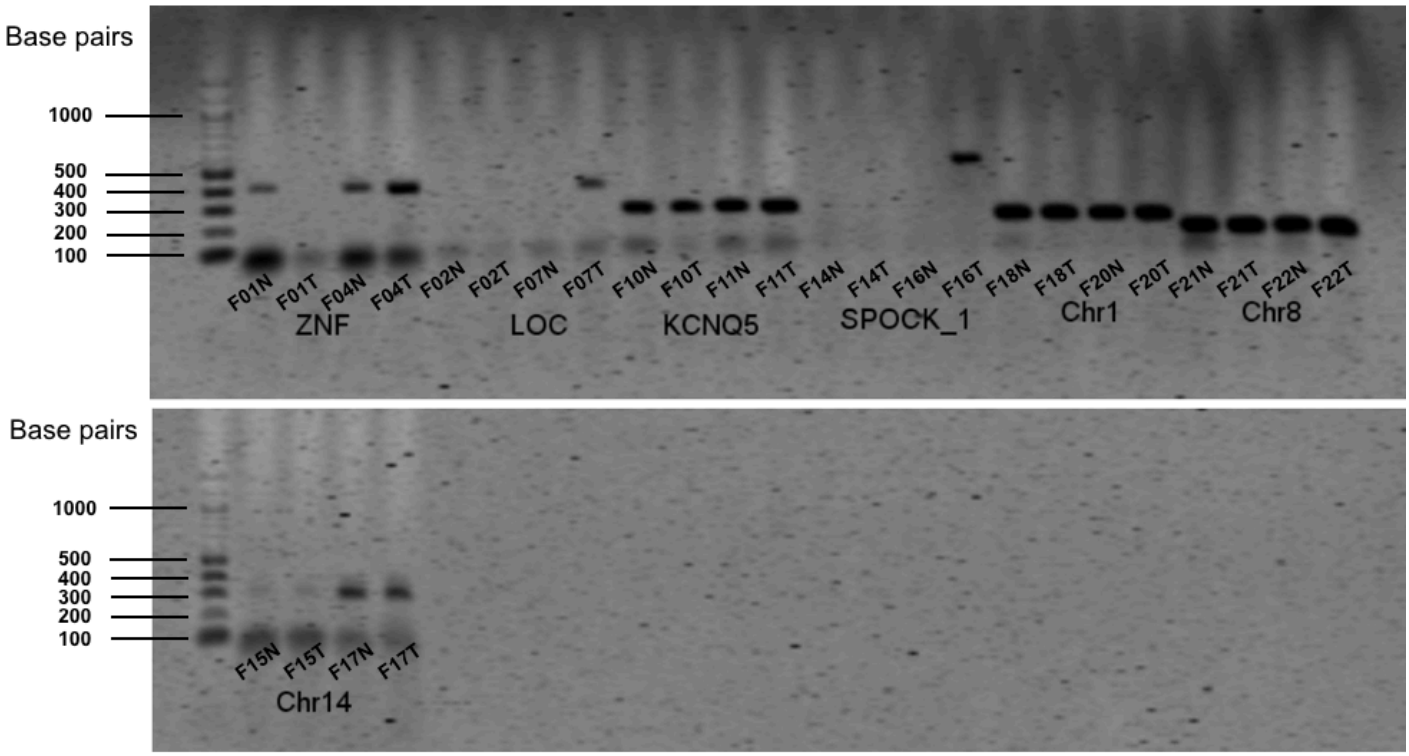


Figure 10.1: Gel electrophoresis image demonstrating the applicability of pyrosequencing assays to FFPE tissue samples. Amplification of four FFPE DNA samples was attempted per assay. KCNQ5, Chr1 and Chr8 each displayed strong bands following amplification of all samples assessed.

10.3 Pyrosequencing

KCNQ5, Chr1 and Chr8 each displayed strong bands following amplification of all samples assessed. *SPOCK*_1 failed to produce amplicons from any samples. Bands consistent with amplification of *ZNF677*, LOC28399 and Chr14 were apparent although these were weaker than those of KCNQ5, Chr1 and Chr8. *KCNQ5*, Chr1, Chr8 and *ZNF677* were therefore selected for further analysis by pyrosequencing.

Pyrograms of *KCNQ5* and *ZNF677* failed to reach sufficient peak height for reliable interpretation. Chr 1 and Chr8 were assessed successfully and correlated strongly with average DNA methylation values of the historical cohort (r=0.98, Figure 10.2).

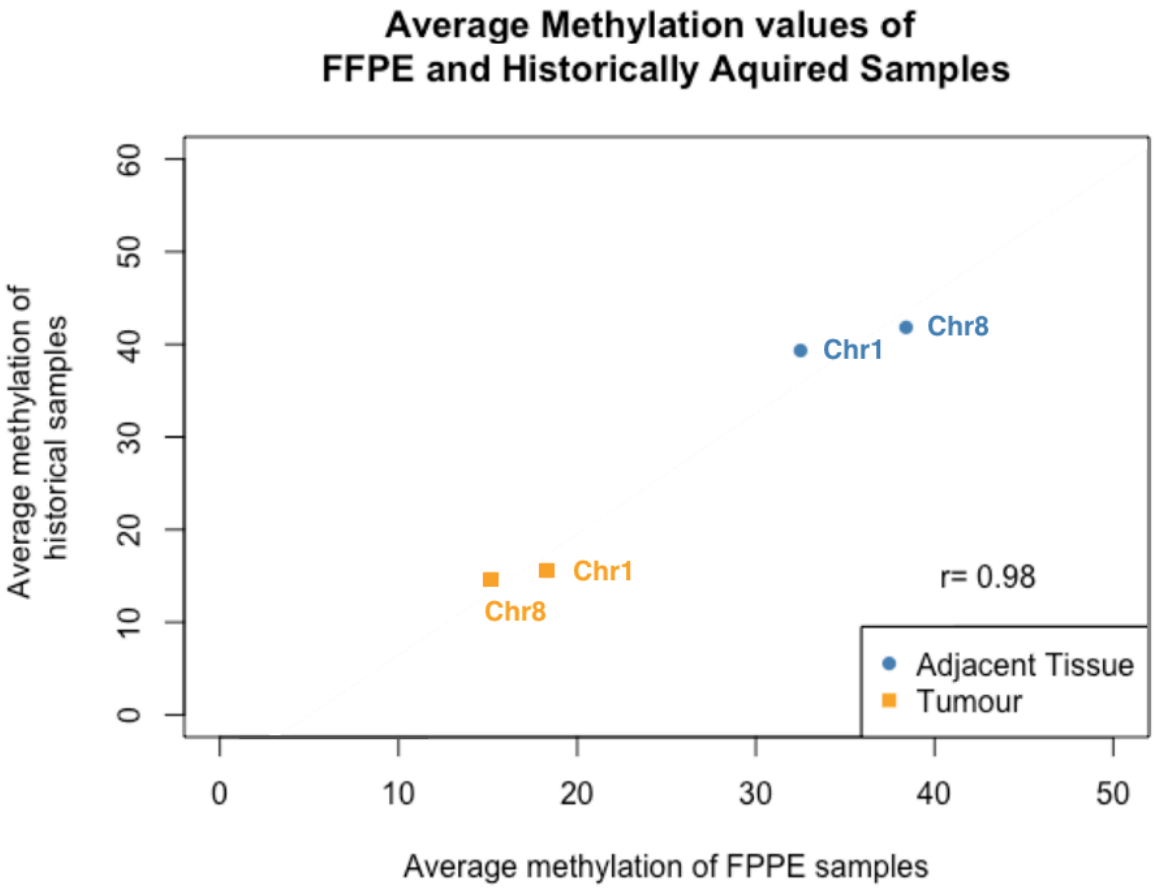


Figure 10.2: Correlation between average methylation values of historical and FFPE tissue samples at Chr1 and Chr8. These findings demonstrate that where bisulphite pyrosequencing assays are able to successfully assess DNA methylation in DNA obtained from FFPE samples, the data obtained is very comparable to that of fresh frozen tissue.

These findings demonstrate that where bisulphite pyrosequencing assays are able to successfully assess DNA methylation in DNA obtained from FFPE samples, the data obtained is very comparable to that of fresh frozen tissue. The success of PCR amplification was correlated with amplicon size (Table 5.2) with sufficient product for bisulphite pyrosequencing produced in the small Chr 1 (209bp) and Chr 8 (154bp) assays only. *SPOCK*_1 was the largest assay (542bp) and failed to amplify DNA from FFPE samples at all. Future studies should therefore aim to produce amplicons of the smallest possible size if work with DNA from FFPE tissues is anticipated.

11 Discussion

11.1 Summary of findings

The primary purpose of this study was to validate DNA methylation differences between rectal tumour and normal mucosa as determined by our preliminary EWAS. All DMPs selected for validation demonstrated strong correlations between bisulphite pyrosequencing and Illumina 450k methylation values (r=0.87-0.97). Replication in an independent CRC cohort (n=68) mirrored findings of the initial cohort with significant levels of DNA methylation noted at all CpG sites assessed (P = <1.95e-17). These findings support our hypothesis that differences observed reflect true biological differences. Global DNA hypomethylation was observed in RT (54.6%) compared to AM (63.5%, P = 0.021) using a modified luminometric methylation assay (LUMA). We explored potential clinicopathological associations with DNA methylation in the replication cohort, however no significant relationships were detected.

11.2 DNA methylation data obtained from 450k genome-wide analysis reflect true biological differences between tumour and adjacent tissue

Strong correlations were observed between pyrosequencing and Illumina 450k array DNA methylation values of all DMPs investigated. This indicates that differences observed reflect true biological differences, as opposed to technical artefacts, between tumour and adjacent mucosa in the RIST sample cohort. This conclusion is further supported by the successful replication in our second independent cohort of 68 CRC patients. Extrapolation of these findings enables us to consider the DNA methylation data obtained for all 176 DMPs to be true. These DMPs may play important biological roles in the pathogenesis of rectal cancer and warrant further investigation in future studies (see Section 11.3).

Replication in an independent CRC cohort (n=68) mirrored findings of the initial cohort with significant levels of DNA methylation noted at all CpG sites assessed (P = <1.95e-17). This suggests that these differences are

generalisable to CRCs more generally and as opposed to the initial RIST cohort only.

Other studies have utilised the Illumina 450k array to identify between 5929-18568 DMPs in CRC tissue samples.(78, 81, 82) These studies however, have employed less stringent bioinformatic approaches and did not correct P values for multiple tests. In contrast, the 176 DMPs identified in our study each achieved significance levels below 1.0E-07.

In addition to the DMP of interest, pyrosequencing enables the assessment of neighbouring CpG sites. In all cases, DNA methylation levels of all CpGs within a given amplicon remained relatively consistent throughout. These DNA methylation values are therefore likely to represent those of the surrounding genome.

The sparc/osteonectin, cwcv and kazal-like domains proteoglycan-1 (*SPOCK_1*) gene, also known as testican-1, encodes a matricellular glycoprotein belonging to a calcium binding proteoglycan family.(160) Other members of this family include *SPARC*, testican-2 and -3 and are implicated in cell proliferation, adhesion, and migration.(161) Our preliminary genome-wide analysis identified two significant DMPs within the 5' untranslated region (UTR) of the *SPOCK_1* gene (cg24847829, P = 6.29E-09; cg14650610, P = 7.30E-10). Due to their close proximity, both of these DMPs were analysed by a single pyrosequencing assay in this study. Significantly differential methylation was observed in both validation and replication cohorts.

Zinc finger protein (*ZNF*) 677 is located at the chromosomal region 19q13. Two DMPs were identified within the 5' UTR of *ZNF677* within our genome-wide analysis (cg24685755, P = 1.39E-08; cg18335068, P = 7.99E-08). It belongs to a diverse family of proteins defined by their requirement for stabilisation by binding to at least one zinc ion.(162) The individual functions of ZNF proteins are diverse but include DNA recognition, RNA packaging, transcriptional activation, regulation of apoptosis, protein folding and lipid binding.(162) DMPs within *ZNF334*, *ZNF625* and *ZNF132* genes were also

highlighted in our genome-wide study and have been previously implicated in CRC, breast and prostate cancers.(80, 163, 164)

Two DMPs (cg05447008, P = 9.87E-9; cg04377145, P = 2.72E-8) were identified within 1500 nucleotides of the transcriptional start site (TSS) of *Potassium voltage-gated channel, KQT-like subfamily, member 5* (*KCNQ5*) following our initial genome-wide analysis. Potassium channels are the most abundant ion channel in eukaryotic cells. Multiple subtypes, each in possession of subtly unique biological functions, exist and are expressed throughout the body. In mammals, five *KCNQ* genes encode voltage-gated potassium channels Kv7.1–Kv7.5 (or simply KCNQ1–5). Other members of the *KCNQ* family have been associated with multiple congenital abnormalities including long QT syndrome, deafness and epilepsy.(165-167) However, little is currently known about the biological implications of *KCNQ5* mutation in human disease.(168)

A single DMP (cg03576469) was highlighted within 200 nucleotides of the coiled-coil domain-containing protein 8 (*CCDC8*) TSS (P = 6.83E-9). *CCDC8* has multiple biological functions and has been implicated in the regulation of gene expression, cell division, and membrane fusion.(169) *CCDC8* also acts as a cofactor for P53-mediated apoptosis following DNA damage and therefore acts as a tumour suppressor.(170)

Cg26034516, situated within the body of LOC283999, was hypermethylated in our genome wide study (P = 6.48E-9). The function of LOC283999 is currently unspecified and to our knowledge, has not been investigated in the context of human disease or cancer.

Likewise, the hypomethylated DMPs investigated in this study were not annotated to any individual gene. They were included however, in order to ensure that both hypermethylated and hypomethylated DMPs could be validated by pyrosequencing. Hypomethylation of intergenic regions is a common finding in cancer and contributes to genomic instability. Global DNA hypomethylation of rectal tumour tissue was also demonstrated in this work using the LUMA assay (see Section 11.5 for full discussion).

11.3 Possible biological implications of findings

As discussed, the purpose of the current study was to validate DNA methylation differences between tumour and normal tissue as nominated by our preliminary EWAS. Mechanistic experiments were not conducted. As a result, we currently remain unable to comment on the biological implications of the DNA methylation changes hightlighted in our work. Other transcriptomic and proteomic analyses conducted by other groups however, may provide some insight.

Overexpression of *SPOCK1* has been demonstrated to occur in multiple forms of cancer including gallbladder, gastric, breast and colorectal.(160, 171-173) The mechanism by which *SPOCK1* expression is increased, however is poorly understood. To date, interactions with chromodomain helicase/adenosine triphosphatase DNA binding protein 1-like (CHD1L), microRNA-129-5p and transforming growth factor beta (TGF-β) have all been implicated. (160, 174, 175) DNA hypermethylation, as in our study, may represent an additional mechanism of SPOCK1 activation.

Shu et al. proposed that *SPOCK1* brings about its effects via two main mechanisms: the induction of EMT (Section 1.5.4) and via the inhibition of the pro-apoptotic phosphoinositide 3-kinases (PI3K)/Akt pathway.(161) These findings are supported by Zhao et al. who investigated the role of *SPOCK1* in CRC cell lines. (171) The P13K/Akt signalling pathway plays an important role in regulating cancer cell growth and proliferation, and has been previously associated with CRC.(176). More recently, activation of the Wnt/β-catenin pathway by *SPOCK1*, has also been demonstrated in glioma cells.(177)

In cancer, the gene acts as an oncogene, with multiple studies reporting reduced cell line viability, invasiveness and propagation following gene silencing.(160, 161, 171) *SPOCK1* knockdown models have demonstrated reduced growth, migration and invasion in CRC cell lines.(171) Subsequent western blot analysis of EMT-related markers, displayed significantly

increased levels of E-cadherin in knockout models— loss of which is a well-established feature of EMT. (178, 179) Loss of this protein weakens intracellular adhesion and consequently increases mobility.

To our knowledge, only one study has described the impact of *ZNF677* expression in cancer.(180) Heller et al. analysed publically available databases in the context of non-small cell lung cancer (NSCLC) and identified tumour specific downregulation of *ZNF677* ($P = < 0.0001$). Further analysis recognised DNA hypermethylation as the cause of reduced expression, with resultant transcriptional effects in multiple growth and interferon signalling pathways.

In a study by Ashkorab et al., promoter region methylation of *KCNQ5* was highlighted to occur consistently within colorectal cancer tissue samples. (181) However, the study did not seek evaluate the effect of this finding further.

Hypermethylation of *CCDC8* has been associated with reduced gene expression in both breast and renal cell carcinomas.(182, 183) Reduced expression of the gene has also been identified in multiple other tumour types including hepatocellular and lung cancer.(184, 185)

Some DMPs identified in our initial EWAS, such as *RNF175*, *PRKAR1B* and *LONRF2*, were not amenable to bisulphite pyrosequencing due to limitations of primer design. Single nucleotide polymorphisms (SNPs) of *PRKAR1B* have previously been implicated in renal cell and non-small cell lung carcinoma, however to our knowledge, have not yet been reported in CRC.(186, 187) In contrast *LONRF2* hypermethylation has been reported to result in reduced gene transcription in rectal adenocarcinoma by Hua et al.(103) The biological and clinical implications of changes in these genes however, have not been identified.

11.4 Clinicopathological implications of DNA methylation

The clinical utility of epigenetic biomarkers depends upon their ability to stratify patients according to clinicopathological features. This concept is of particular importance in the fields of personalised medicine and treatment selection. The potential of epigenetic biomarkers to predict response to neoadjuvant chemoradiotherapy are discussed in Section 11.7 and form the basis of our future work.

Other studies have identified multiple associations between DNA methylation and clinicopathological features. Of these, only a limited number of associations have been reported with the genes investigated in this study. These relate predominantly to overall survival, nodal and distant metastasis.(161, 172, 180, 184) In our study, only two patients had distant metastasis and so comparison was not possible. Associations with nodal involvement were not reproduced in our cohort and overall survival was not assessed. Larger sample sizes, with greater numbers of patients within each clinical subcategory, may have revealed more significant findings.

Importantly, these analyses were very explorative in nature and were not factored into the initial study design. The failure to identify additional associations may therefore be rooted in our initial method of DMP selection (i.e. according to statistical significance). DMPs of the highest statistical significance, are by nature the most likely to occur in all samples. As a result, these changes are expected to occur at an early stage of tumourgenesis and are unlikely to differ considerably between clinicopathological subgroups. Therefore, although this method of selection was well suited to the purpose of validation and replication, other selection methods, such as consideration of biological function, may have yielded findings of greater clinical relevance.

Due to these limitations, it is important to note that failure to identify clinicopathological associations in our study, does not provide definitive evidence that true effects do not exist.

Indeed, animal studies using xenografted *SPOCK1* silenced cancer cells have reported smaller tumour volumes and fewer pulmonary metastases than those with unrestricted *SPOCK1* expression in the context of gallbladder and prostate cancer respectively.(161, 188) In our study, *SPOCK1* hypermethylation did not correlate significantly with any clinicopathological tumour features assessed. In contrast, Zhang et al. investigated the implications of *SPOCK1* gene expression in CRC (n=84).(172) They reported that increased *SPOCK1* expression was associated with increased tumour size (P = 0.020) and TNM status (P = 0.012), but not age, gender, location or lymph node metastasis.(172) In gallbladder cancer, Shu et al. reported that increased expression was significantly associated with histological differentiation (P = 0.012), lymph node metastasis (P = <0.001) and shorter overall survival time (log rank, 11.301; P = 0.001).(161)

Clinical outcome data of 97 non-small cell lung cancer patients also indicated reduced overall survival in the presence of high *ZNF677* methylation levels (P = 0.013). Tumour specific downregulation of *ZNF677* was also identified in other solid tumours including colorectal (P = 1.5E-7), breast (P = 8.2E-42) and renal cancer (P = 1.0E-10).(180)

In non-small cell lung cancer, Jiang et al. identified that reduced *CCDC8* expression was associated with tumour differentiation (P = 0.039), TNM stage (P = 0.009), lymph node metastasis (P = 0.038), and overall survival (P = .043).(184) The study also identified a direct relationship between *CCDC8* and E-cadherin expression— which may result from the inhibition of pathway signalling molecules such as p-P38, p-IκBα and Snail. Reduced expression therefore, may promote EMT and tumour invasiveness. These findings are supported by Pangeni et al. who also demonstrated the presence of *CCDC8* hypermethyaltion in both primary and secondary metastatic breast tumours— indicating that the change occurs at an early stage of metastatic evolution.(182)

11.5 Rectal cancer tissues display global DNA hypomethylation

The LUMA assay was initially described by Karimi et al. in 2006.(151) The utilisation of methylation dependent and independent isoschizomers enables the estimation of global DNA methylation. Since its initial description, multiple modifications to the LUMA assay have been proposed (see 4.10.1). These modifications primarily aim to overcome the major limitation of the LUMA assay— the requirement for high quality, unfragmented DNA. In this study, the substitution of EcoRI for MunI, and modification of the pyrosequencing dispensation order, as described by Lisanti et al. and Bjorrnson et al., were adopted.(153, 155) The combination of these modifications resulted in the highest predicted value of global DNA methylation. As nonspecific fragmentation of DNA results in the underestimation of global DNA methylation, a higher percentage of methylation was attributed to greater compensation for fragmentation.

In this study, sufficient DNA for LUMA analysis was available for only ten of the 15 RIST sample pairs. However, global DNA methylation was significantly reduced in tumour samples when compared to adjacent tissue. These findings of global DNA hypomethylation, in conjunction with hypermethylation of specific genes, reflect two key epigenetic mechanisms by which normal cells become malignant.

Due to the selective analysis of CCGG sequences, the LUMA assay does not assess all CpGs in the genome. The CCGG sequence is more highly enriched within gene promoters and CpG islands (11.7 and 12.9%) than repetitive elements and other unique sequences (4.1 and 3.9%).(156) As the latter regions are usually the principal sites of DNA hypomethylation in cancer, the LUMA assay may not fully reflect the true methylation changes inherent to malignancy.

Other methods of global DNA methylation quantification are available and assess different regions of the genome.(189) As DNA hypomethylation occurs predominantly in intergenic regions such as long interspersed nucleotide

element-1 (LINE-1, see Section 1.5.1), specific LINE-1 assays are likely to detect greater differences than those observed in our study.(156) In a study by Knothe et al., although global DNA methylation values determined by LINE-1 and LUMA were correlated, tissue-specific biases were observed between the two methods.(156)

11.6 Limitations of the study

11.6.1 RIST samples

We acknowledge multiple limitations of this study. The relatively small number of samples within the initial RIST cohort, were obtained from a single centre in the South West of England and therefore may not be representative of changes seen in the population more generally. Samples obtained by biopsy were too small for formal histological assessment and tumour characteristics were therefore obtained from separate samples. Finally, the Illumina Human Methylation 450 BeadChip array, used in the preliminary EWAS has its own limitations. Although the array it interrogates 482422 cytosine sites, these represent only around 1.7% of all CpG sites in the human genome. This is confounded by the fact that CpG sites assessed are predetermined and as a result, many important sites may have been missed—some of which may configure important biological properties.

11.6.2 Tumour heterogeneity

Colorectal cancer is a highly heterogeneous disease with significant variation in tumour features both between and within individual tumours. As a result, the amount of tumour present at any given site is variable. Some studies have utilised minimum tumour load for inclusion of tissues. Few however, provide rationale for this decision. As DNA methylation differences would be expected to increase with increasing tumour cellularity, we expected to observe a strong correlation between the two parameters. However, this relationship was not evident in our analysis. Furthermore, high and low DNA methylation values were observed at both extremes of tumour cellularity. The employment of a minimum threshold in our study therefore appeared arbitrary.

In a study by Tournier et al., a maximal deviation of 3.6% methylation was observed despite 20% variation in tumour cellularity between samples of the same tumour.(190) Given the large DNA methylation differences between tumour and non-tumour tissue in our cohorts, such variations are unlikely to be of significant impact.

Other techniques to enrich for tumour cell content and their impact on DNA methylation measurements have been described. Ishara et al. compared the global DNA methylation levels of tumour tissues prepared by macro dissection (i.e. as in our study) and laser capture micro dissection (LCM)—a method which enables isolation of tumour cells from a sample.(191) The study found no significant impact of contamination by non-malignant cells on global DNA methylation. These findings are supported by Leong et al. who found no advantage of LCM over manual dissection in the assessment of *APC* gene methylation.(192)

These findings are of relevance as samples collected from the initial RIST cohort were too small for histological assessment and therefore analysed without quantification of tumour load. RIST tissue samples were obtained primarily by tissue biopsy at endoscopy or at the time of surgery. Biopsies are easy and quick to perform. However due to tumour heterogeneity, biopsies taken from a tumour may not be completely representative of the tumour as a whole. As discussed, it is likely that the DMPs investigated in our study occur early in tumourgenesis. As a result, such changes may be present in apparently normal cells prior to malignant transformation.

Furthermore, colorectal tumours are highly heterogeneous in their molecular characteristics.(193, 194) As a result, it is plausible that the tumour load of immediately adjacent tissue may not reflect that of the tissue used for DNA extraction anyway. Ooki et al. investigated the extent by which methylation status is influenced by tumour heterogeneity in relation to the *HOPX*-β gene.(195) They analysed superficial, intermediate and deep tumour tissue and reported no significant difference in DNA methylation status between the sections.

11.6.3 Possible implications of 'field change'

Field change describes the phenomenon by which seemingly non-malignant cells develop underlying pro-oncogenic molecular, epigenetic and genetic changes consistent with neoplasia.(196) Multiple models to explain the underlying mechanism of field change have been proposed.(197) Each of these models stem from the notion that similar environmental exposures within an organ, predispose the surrounding tissues to undergo the same neoplastic transformation as seen within the primary tumour. Subsequent genetic and epigenetic changes subsequently lead to the clonal expansion of cellular populations with changes advantageous to growth and survival.

The impact of field cancerisation in rectal cancer remains unclear. In a genome-wide study conducted by Naumov et al., little overall variation was seen in the DNA methylation signature of tissue taken adjacent to rectal carcinoma and that of healthy controls.(78) Other studies have identified specific genes implicated in field change including *MGMT, P14*^{ARF} and *EVL/miR-342*.(198, 199) In a study by Grady et al., the adjacent mucosa (10cm from tumour) of patients with CRC displayed increased methylation of *EVL/miR-342* in comparison to healthy controls.(199) These findings imply a large field of change. In contrast, Shen et al. reported greater *MGMT* methylation in adjacent tissue samples taken 1cm from the tumour than those taken at 10cm.(200) It is therefore likely that different genes are involved in field change to differing extents.

In our study, the distance between adjacent tissue and tumour was decided at the discretion of the surgeon performing the procedure. No attempt was made to standardise the biopsy method and the distance was not measured in either cohort. As a result, we are unable to comment on whether our DMPs were influenced by field change. The phenomenon may be exemplified by greater variance between DNA methylation levels of the non-tumour samples, although this was not assessed in the current study. However, as genes involved in field change would display smaller differences between tumour

and adjacent tissue, the impact of this phenomenon is more likely to have been influential in the initial EWAS study than subsequent pyrosequencing assays.

11.7 Future work: Risk Stratification in Rectal Cancer (RIST)

11.7.1 Overview of RIST

The work described in this thesis forms part of a larger study, which aims to identify epigenetic biomarkers to predict response to neoadjuvant chemoradiotherapy (nCRT).

Surgery to remove the rectum has traditionally been considered the only treatment option to offer long term cure in locally advanced rectal cancer.(201) However, this carries significant risks of complications, poor long term function and reduced quality of life.(202-204)

The use of nCRT is intended to reduce tumour bulk prior to surgery and increase the chances of complete excision. However, the response to nCRT is highly variable.(205)

The ability to predict who will and will not respond to CRT is of crucial importance to improving patient treatment and care. The project aims to explore epigenetic and transcriptional changes that may be associated with outcomes from rectal cancer treatment. This may enable us to identify potential methylomic biomarkers for cancer prognosis and prediction of response to nCRT. These biomarkers would then be evaluated in future larger prospective studies.

The RIST study aims to identify novel epigenetic changes that can be used individually, or together in a panel, to predict the response of patients with locally advanced rectal cancer to nCRT.

11.7.2 Lessons learnt from this study

In addition to the matched tissue samples analysed in this study (RIST001-RIST016), additional samples from a total of 45 patients have now been obtained. Through this work, we identified that fewer patients than expected were treated with nCRT at our centre. In addition, coordination of all team members to ensure effective patient recruitment and tissue collection was not always possible. As a result, our future studies may benefit from a multicentre approach to tissue collection in order to ensure that adequate sample sizes can be obtained.

Through our work, our team has gained greater understanding and experience with methodological processed employed in this study. The 450k genome-wide analysis with pyrosequencing as a validation method have proven to be suitable assessment methods of paired tumour samples. This study successfully compared tumour with adjacent non-malignant mucosa. In order to identify specific DMPs associated with response to nCRT, the RIST study should make direct comparisons of patients with successful and unsuccessful treatment responses.

As discussed in Section 11.6, tumour DNA methylation was not significantly influenced by tumour load. This is important as tissues in the RIST cohort were obtained predominantly from tissue biopsies and were therefore too small for histological assessment. The lack of obvious relationship between tumour cellularity and DNA methylation levels support the notion that histological assessment is of limited value and therefore not required.

In this work, no significant clinicopathological associations were identified with the nine DMPs investigated. The DMPs selected for validation in this work were chosen according to their statistical significance from data obtained by comparison of tumour and adjacent tissue. As a result, they are likely to be common to all tumour samples and probably occur early in tumorigenesis. As a result, they are unlikely to impart any significant influence on tumour characteristics or behaviour. Other studies have selected DMPs on the basis of known biological function. This method is limited however, in its

requirement for an idea of which genes are likely to be implicated. It would therefore be unable to identify novel markers in the absence of obvious biological mechanisms. Direct comparison of epigenome-wide DNA methylation data obtained from responders and non-responders, would represent an unbiased approach to this problem.

12 Conclusion

This study, in conjunction with our previously conducted EWAS, has successfully identified and validated a novel methylomic signature of rectal cancer. Successful validation of EWAS data enabled us to conclude that its findings reflect true biological differences between cancer and normal tissue. Although no clinicopathological correlations were observed with the DMPs investigated, others may represent potential targets in the diagnosis, monitoring and risk stratification of rectal cancer. Future work will use the methodologies employed to investigate the relationship between DNA methylation and response to neoadjuvant chemotherapy.

Appendices

1. DMPs identified by epigenome-wide association study
2. RIST pyrosequencing and Ilumina 450k DNA methylation data
3. Pyrosequencing DNA methylation data of historical cohort
4. Clincopathological features of historical cohort

Appendix 1: DMPs identified by epigenome-wide association study

CpG ID	Array P value	Array Δ/%	Chr	Position	Annotated gene
cg02647878	3.66E-10	-0.45	4	154681197	*RNF175*
cg14650610	7.30E-10	-0.48	5	136834492	*SPOCK1*
cg13001868	1.58E-09	-0.45	17	43339223	*C17orf46*
cg13895235	2.22E-09	-0.57	7	752292	*PRKAR1B*
cg18538668	3.79E-09	0.20	14	103839038	
cg03061682	4.83E-09	-0.42	15	28352098	
cg24847829	6.29E-09	-0.40	5	136834464	*SPOCK1*
cg26034516	6.48E-09	-0.35	17	76228121	*LOC283999*
cg03576469	6.83E-09	-0.29	19	46917061	*CCDC8*
cg18601167	8.02E-09	-0.56	7	752286	*PRKAR1B*
cg05447008	9.87E-09	-0.40	6	73331114	*KCNQ5*
cg13356896	1.15E-08	-0.40	2	198650987	*BOLL*
cg09129067	1.29E-08	0.35	8	103750904	
cg23977631	1.30E-08	-0.46	2	100938799	*LONRF2*
cg26238800	1.38E-08	-0.37	20	45142206	*ZNF334*
cg24685755	1.39E-08	-0.35	19	53758031	*ZNF677*
cg24820783	1.41E-08	-0.25	10	26504969	*GAD2*
cg25480336	1.50E-08	-0.39	20	50720908	*ZFP64*
cg16964348	1.65E-08	-0.39	7	24323799	*NPY*
cg11328303	1.94E-08	-0.27	10	26505440	*GAD2*
cg09734791	2.06E-08	-0.46	8	72756155	*MSC*
cg10224098	2.26E-08	-0.40	1	44873229	*RNF220*
cg04921989	2.32E-08	-0.37	2	132183100	
cg26314722	2.36E-08	0.24	1	234867300	
cg04504205	2.42E-08	0.32	20	45946429	*ZMYND8*
cg11220565	2.48E-08	-0.36	20	47934802	
cg18324583	2.48E-08	0.22	5	142975083	
cg19752627	2.51E-08	-0.39	7	98467380	*TMEM130*
cg25884711	2.52E-08	-0.42	7	24323840	*NPY*
cg03470088	2.52E-08	-0.03	1	24513939	*IL28RA*
cg06952671	2.55E-08	-0.47	2	182322268	*ITGA4*
cg04377145	2.72E-08	-0.33	6	73331191	*KCNQ5*
cg14168530	2.84E-08	-0.51	2	45155991	
cg00859129	2.87E-08	0.22	1	109422184	*GPSM2*
cg26958524	2.90E-08	-0.34	16	86613067	*FOXL1*
cg25024074	2.94E-08	-0.31	2	182322501	*ITGA4*
cg26328335	3.26E-08	-0.45	12	50354840	*AQP5*
cg22434409	3.29E-08	-0.41	4	21950722	*KCNIP4*
cg26020069	3.29E-08	0.27	6	52382441	*TRAM2*
cg08266366	3.38E-08	-0.38	12	50354998	*AQP5*
cg18607529	3.47E-08	-0.41	7	50343869	*IKZF1*

cg06319475	3.67E-08	-0.38	8	145105829	
cg07589773	3.80E-08	-0.37	7	50343883	*IKZF1*
cg03020208	3.85E-08	-0.30	12	50354962	*AQP5*
cg11751707	3.85E-08	-0.27	2	38302587	*CYP1B1*
cg09220050	3.88E-08	0.03	20	48770642	*TMEM189*
cg20415809	3.92E-08	-0.39	2	182321855	*ITGA4*
cg21938148	4.01E-08	-0.48	13	110958977	*COL4A1;COL4A2*
cg13405887	4.37E-08	-0.42	9	132382812	*C9orf50*
cg25340966	4.46E-08	-0.34	1	119532195	*TBX15*
cg17200768	4.62E-08	-0.39	13	28503373	
cg26415547	4.82E-08	-0.32	12	66583048	*IRAK3*
cg07921384	4.84E-08	-0.33	2	220299740	*SPEG*
cg10013343	4.86E-08	-0.50	13	29106503	
cg12628196	4.87E-08	-0.40	7	127672458	*SND1;LRRC4*
cg06072021	4.99E-08	-0.52	11	128564106	*FLI1*
cg16674351	5.03E-08	-0.35	1	121260892	*LOC647121*
cg24190603	5.05E-08	-0.36	6	84418433	*SNAP91*
cg17170568	5.05E-08	-0.18	7	156433406	*C7orf13;RNF32*
cg15336765	5.15E-08	-0.43	12	50355307	*AQP5*
cg25223771	5.45E-08	-0.38	8	145105503	
cg00741836	5.50E-08	-0.45	20	53092233	*DOK5*
cg02155398	5.53E-08	-0.36	2	45160490	
cg25680916	5.76E-08	0.03	3	53916015	*ACTR8*
cg24924779	5.98E-08	-0.38	20	49639998	*KCNG1*
cg25773267	6.14E-08	-0.36	20	61992187	*CHRNA4*
cg17393267	6.15E-08	-0.29	3	192127356	*FGF12*
cg12441126	6.21E-08	-0.41	7	751962	*PRKAR1B*
cg13267264	6.29E-08	-0.43	8	70983600	*PRDM14*
cg11901272	6.34E-08	-0.38	6	29760447	*HCG4*
cg11947981	6.39E-08	-0.39	2	182322749	*ITGA4*
cg02742906	6.39E-08	-0.34	13	112758625	
cg26718707	6.46E-08	0.42	10	518370	*DIP2C*
cg20381963	6.48E-08	-0.51	7	752238	*PRKAR1B*
cg20107395	6.60E-08	-0.50	20	53092334	*DOK5*
cg24242823	6.79E-08	-0.37	7	24323675	*NPY*
cg11601252	6.92E-08	-0.35	15	68122139	*LBXCOR1*
cg04366687	7.26E-08	-0.35	8	145107199	*OPLAH*
cg18355902	7.28E-08	-0.40	4	154681128	*RNF175*
cg08206318	7.33E-08	-0.30	5	134363637	*PITX1*
cg13554086	7.39E-08	-0.39	5	76507100	*PDE8B*
cg25645268	7.81E-08	-0.31	4	154710598	*SFRP2*
cg17101450	7.85E-08	-0.38	10	102900365	
cg00321614	7.88E-08	0.17	5	172856932	
cg14215472	7.91E-08	-0.35	17	27940404	*ANKRD13B*
cg18335068	7.99E-08	-0.32	19	53757910	*ZNF677*

cg09296001	8.21E-08	-0.52	7	127672564	*SND1*
cg03424342	8.35E-08	-0.28	3	120169783	*FSTL1*
cg26593267	8.47E-08	-0.34	13	113764871	*F7*
cg10120816	8.61E-08	-0.26	6	99296305	
cg09871471	8.62E-08	-0.41	1	121260900	*LOC647121*
cg03147907	9.20E-08	0.34	3	62926690	
cg07719492	9.24E-08	-0.32	8	70983348	*PRDM14*
cg10928466	9.24E-08	-0.11	19	11353961	*DOCK6*
cg21995919	9.27E-08	-0.41	2	182322279	*ITGA4*
cg10770742	9.31E-08	-0.37	7	151107285	*WDR86*
cg23383871	9.31E-08	-0.37	20	47934987	
cg04023150	9.33E-08	-0.45	1	44873064	*RNF220*
cg17228900	9.42E-08	-0.45	6	391764	*IRF4*
cg27532621	9.61E-08	-0.23	1	164593763	*PBX1*
cg10065823	1.00E-07	-0.30	9	96108467	*C9orf129*
cg11573679	1.03E-07	-0.48	2	68546467	*CNRIP1*
cg23572908	1.05E-07	-0.37	7	158937969	*VIPR2*
cg01440841	1.06E-07	-0.35	4	154681066	*RNF175*
cg05946309	1.07E-07	0.21	16	85926085	
cg05288172	1.07E-07	0.27	8	103751006	
cg18918321	1.08E-07	-0.26	8	41424524	
cg16437728	1.09E-07	-0.36	11	7273046	*SYT9*
cg02177231	1.09E-07	-0.33	1	119529930	*TBX15*
cg08957069	1.11E-07	-0.32	6	28743700	
cg23934404	1.13E-07	-0.42	13	112758491	
cg23933289	1.14E-07	-0.23	1	178998656	*FAM20B*
cg11338643	1.16E-07	-0.41	6	166580983	*T*
cg08750504	1.20E-07	-0.39	2	172946193	
cg09073398	1.20E-07	-0.39	5	168727762	*SLIT3*
cg14337134	1.21E-07	-0.18	7	102920323	*DPY19L2P2*
cg21232488	1.23E-07	0.28	6	30079203	*TRIM31*
cg16504626	1.25E-07	-0.38	8	57070013	
cg01618245	1.29E-07	0.24	20	61990279	*CHRNA4*
cg14015706	1.30E-07	-0.50	9	132382433	*C9orf50*
cg14443519	1.30E-07	-0.43	6	29760410	*HCG4*
cg27141850	1.30E-07	-0.38	2	20869434	*GDF7*
cg24171907	1.32E-07	-0.45	2	68546579	*CNRIP1*
cg20652954	1.33E-07	0.20	20	61716293	
cg10210594	1.34E-07	-0.33	1	208132787	
cg12433277	1.34E-07	-0.31	7	151106990	*WDR86*
cg02700626	1.35E-07	-0.24	11	64739320	
cg20295442	1.37E-07	-0.48	8	67344665	*ADHFE1*
cg06401021	1.37E-07	-0.42	6	55443868	*HMGCLL1*
cg21647227	1.37E-07	-0.36	1	119527111	*TBX15*
cg08332074	1.37E-07	-0.31	16	51189941	

cg23641267	1.39E-07	-0.25	11	58343791	*LPXN*
cg23092823	1.39E-07	-0.25	1	53528612	*PODN*
cg14595003	1.41E-07	-0.38	3	129694156	*TRH*
cg27341128	1.41E-07	-0.33	20	53092259	*DOK5*
cg25771271	1.41E-07	-0.32	1	119550191	
cg25189564	1.43E-07	-0.29	7	158938051	*VIPR2*
cg24645214	1.45E-07	-0.30	8	54789978	*RGS20*
cg07878486	1.46E-07	-0.33	19	58951885	*ZNF132*
cg18303242	1.48E-07	0.37	17	25879250	*KSR1*
cg17892556	1.49E-07	-0.47	19	12267464	*ZNF625*
cg12619536	1.49E-07	-0.42	1	108508067	*VAV3*
cg04415599	1.49E-07	-0.24	19	37464508	
cg08516516	1.50E-07	-0.30	5	115152492	*CDO1*
cg02065637	1.51E-07	-0.36	20	61809035	*MIR124-3*
cg08569799	1.52E-07	-0.29	5	1886828	
cg17226446	1.53E-07	-0.14	4	154408845	*KIAA0922*
cg19991022	1.56E-07	-0.55	20	53091929	*DOK5*
cg25060829	1.56E-07	-0.31	6	28367571	*ZSCAN12*
cg13850380	1.60E-07	-0.32	1	1475143	*C1orf70*
cg27200446	1.62E-07	-0.57	6	41606439	*MDFI*
cg15745900	1.62E-07	-0.38	8	68864549	*PREX2*
cg20912169	1.64E-07	-0.48	8	67344720	*ADHFE1*
cg07790085	1.64E-07	0.25	13	29597447	*MTUS2*
cg00250422	1.65E-07	-0.42	15	28352347	
cg09632907	1.68E-07	-0.24	4	54969963	
cg03064067	1.70E-07	-0.39	12	85306916	*SLC6A15*
cg18435449	1.70E-07	-0.33	19	58095445	*ZIK1*
cg17371081	1.70E-07	-0.28	11	20690957	*NELL1*
cg12868067	1.71E-07	-0.38	12	128752246	*TMEM132C*
cg13389502	1.72E-07	-0.16	17	1961440	*HIC1*
cg06427779	1.73E-07	-0.32	5	54180079	
cg27188703	1.76E-07	-0.33	12	50297581	*FAIM2*
cg11903130	1.82E-07	-0.35	10	26506751	*GAD2*
cg18921980	1.82E-07	-0.28	2	175594943	
cg16306898	1.83E-07	-0.48	1	1475675	*C1orf70*
cg02640612	1.84E-07	-0.28	8	53853444	*NPBWR1*
cg16366473	1.88E-07	-0.41	3	192126849	*FGF12*
cg22862480	1.91E-07	-0.49	10	7450355	*SFMBT2*
cg14658804	1.92E-07	-0.33	5	168728213	*SLIT3*
cg23575688	1.93E-07	0.21	11	119486443	
cg21013866	1.95E-07	-0.36	14	23834985	*EFS*
cg24663256	1.96E-07	-0.38	4	21950307	*KCNIP4*
cg22663389	1.97E-07	0.35	17	57929274	
cg01046104	1.98E-07	-0.26	19	58095588	*ZIK1*
cg04803843	1.99E-07	-0.42	15	28351906	

Appendix 2: RIST pyrosequencing and Ilumina 450k methylation data

	Type	KCNQ5 Pyro	KCNQ5 450K	ZNF677 Pyro	ZNF677 450K	Chr1 Pyro	Chr1 450K	Chr8 Pyro	Chr8 450K	Chr14 Pyro	Chr14 450K	SPOCK (5) Pyro	SPOCK (5) 450K	SPOCK (7) Pyro	SPOCK (7) 450K	LOC Pyro	LOC 450K	CCDC8 Pyro	CCDC8 450K
N001	N	18	34.7	14	37	32	40.9	37	57.8	32	36.1	16	29	10	21.9	19	15.2	30	25
T001	T	37	59	21	63.5	14	12.5	18	18.7	22	25.2	59	61	49	66.1	33	47.3	34	40.2
N002	N	18	29.9	17	50.6	43	42	35	45.7	57	48.8	7	24.2	7	24.1	13	17.4	44	42.1
T002	T	56	70.4	48	81.6	10	13.9	10	12.7	15	22.2	68	72.6	68	79.5	59	50.1	69	75.8
N004	N	23	46.2	17	45.6	39	38	51	66.5	54	50.4	15	38.8	11	22.1	19	35.9	41	42.8
T004	T	77	85.5	50	72.7	15	18.6	6	12.1	11	22.2	57	75.5	59	82.6	33	59.6	50	65.6
N005	N	11	26.6	13	45.6	37	38.3	43	51.6	48	42	14	32.2	11	21.2	14	30.1		36.5
T005	T	43	65.5	82	81.2	15	10.8	4	12.2	7	19	81	86.1	30	88.9	37	58.5		81.2
N006	N	19	35.6	21	43	38	41.2	45	48.9	45	47.9		41.5		31	17	26.1	53	53.3
T006	T	90	87.4	84	88.9	5	11.9	4	10	6	21		88.1		92.8	82	67.4	86	88.4
N007	N	22	26.2	21	48.1	39	37.9	60	69.2	46	51.6	15	34.6	6	27.3	11	21.2		40.6
T007	T	62	67.8	49	74.9	17	18.4	28	44.2	30	35.3	44	55.9	28	48.6	34	53.5		66
N008	N	9	12.3	7	24.3	45	45.6	42	53.5	43	35.8	5	16.4	2	22.5	8	11.4	25	25.4
T008	T	17	63.4	30	61.4	8	13	11	12.5	12	21.1	71	73.4	53	68.2	62	61.9	46	50.4
N010	N	15	23.1	11	42.5	40	40	47	64.4	46	38.1	9	31.8	12	11.5	4	13.3	38	40.5
T010	T	30	38.9	29	71.7	14	12.2	20	28.3	27	29.2	35	50.9	34	46.9	13	42	54	60.2
N011	N	8	18.4	14	35.1	50	47.1	47	55.8	48	39.7	13	12.9	12	7.6	11	13.8	38	21.6
T011	T	39	66.8	34	72.5	17	17.3	15	22.8	22	23	29	61.2	30	62.4	56	56.4	54	61
N012	N	12	24.3	13	44	40	44	54	65.8	52	46.9	7	37.9	9	26.5	11	27.6		51.4
T012	T	89	85.5	87	89.9	10	16.7	4	6.9	15	25.7	86	83.3	58	80.5	75	68.3		85
N013	N	7	29	5	25.6	37	37.6	37	42.6	50	58.6	24	26.9	13	17.8	2	22.3	24	37.5
T013	T	74	69.3	61	79	12	15.6	10	20.6	24	36.3	77	72.9	59	74.4	40	64.8	70	77.7

	Type	KCNQ5 Pyro	KCNQ5 450K	ZNF 677 Pyro	ZNF 677 450K	Chr1 Pyro	Chr1 450K	Chr8 Pyro	Chr8 450K	Chr 14 Pyro	Chr 14 450K	SPOCK (5) Pyro	SPOCK (5) 450K	SPOCK (7) Pyro	SPOCK (7) 450K	LOC Pyro	LOC 450K	CCDC8 Pyro	CCDC8 450K
N014	N	19	32.4	16	44.4	35	37.1	36	52.3	42	44.7	14	46.6	9	45.9	11	27	38	43.4
T014	T	91	85.3	48	83.7	9	16.8	7	17.3	15	26.4	84	81.5	92	86.7	67	66.8	77	77.5
N015	N	15	34.2	11	31.2	52	47.2	26	32.8	57	53.3	12	21.4	8	14.8	8	20.1		35.5
T015	T	65	73.3	57	81.7	17	17.2	6	13.4	25	27.2	31	62.4	41	65.1	43	62.8		66.2
N016	N	7	16.8	7	22.7	34	34.6	39	63	50	51.8	9	14.2	4	10.4	5	10.2		30.8
T016	T	40	56	21	55.4	21	18.4	27	32.8	40	34.3	32	61.2	26	62	22	46.5		48.8

Appendix 3: Pyrosequencing DNA methylation values of historical cohort

ID	ZNF.N	ZNF.T	KCNQ5.N	KCNQ5.T	SPOCK5.N	SPOCK5.T	SPOCK7.N	SPOCK7.T	LOC.N	LOC.T	Chr1.N	Chr1.T	Chr8.N	Chr8.T	Chr14.N	Chr14.T	Chr14.D
3CR	3	28	4	39	4	37	5	40	3	32	29	11	40	22	19	17	-2
4CR	7	67	11	73	12	55	10	54	7	59	41	6	46	7	45	16	-29
6CR	19	35	2	43	NA	NA	NA	NA	7	33	47	13	41	17	45	22	-23
8CR	8	32	6	58	15	41	9	15	6	25	28	15	42	25	46	19	-27
11CR	7	35	6	48	10	14	7	45	9	52	41	7	49	5	49	10	-39
13CR	18	18	8	33	13	33	8	2	8	21	48	10	41	9	42	16	-26
15CR	22	40	14	57	20	49	11	48	11	40	41	8	47	17	40	8	-32
20CR	5	29	7	35	7	34	5	35	4	17	42	14	40	10	49	40	-9
24CR	18	34	16	40	16	25	11	50	15	56	33	10	37	9	52	16	-36
25CR	20	51	20	58	18	45	16	52	12	30	27	12	38	14	42	12	-30
28CR	13	20	5	39	14	42	10	9	8	30	43	19	40	15	41	35	-6
30CR	20	57	16	78	15	77	5	75	12	62	22	22	52	38	36	8	-28
31CR	NA	NA	12	56	13	50	10	48	10	89	31	18	NA	NA	NA	NA	NA
33CR	23	29	22	37	NA	NA	NA	NA	21	21	27	11	34	10	40	21	-19
35CR	11	47	6	39	7	11	6	11	7	38	34	9	36	12	49	15	-34
37CR	8	22	8	15	NA	NA	NA	NA	5	10	46	13	49	32	48	34	-14
38CR	22	64	14	67	13	69	11	69	10	52	38	11	39	7	45	14	-31
39CR	5	64	8	62	8	63	6	61	8	48	37	8	47	6	48	19	-29
40CR	17	56	14	60	13	55	11	71	13	43	40	29	45	13	42	24	-18
41CR	11	29	3	56	12	51	7	53	5	36	33	10	44	11	40	19	-21
42CR	24	37	6	22	10	20	7	15	5	20	39	19	39	12	39	34	-5
43CR	19	59	19	67	16	64	12	63	15	58	35	7	45	10	41	24	-17

ID	ZNF.N	ZNF.T	KCNQ5.N	KCNQ5.T	SPOCK5.N	SPOCK5.T	SPOCK7.N	SPOCK7.T	LOC.N	LOC.T	Chr1.N	Chr1.T	Chr8.N	Chr8.T	Chr14.N	Chr14.T	Chr14.D
44CR	19	79	12	91	13	84	4	82	8	72	NA	NA	39	3	42	49	7
45CR	18	51	10	56	13	58	6	24	9	38	48	12	39	12	39	18	-21
47CR	20	64	22	77	19	3	11	10	12	61	43	16	45	10	40	12	-28
51CR	18	17	NA	NA	14	20	10	10	6	14	36	31	40	37	51	47	-4
52CR	16	65	12	79	11	68	11	72	9	42	41	7	34	4	17	9	-8
56CR	10	49	15	56	17	50	10	49	8	28	40	15	44	18	41	31	-10
57CR	14	15	9	15	11	14	9	4	5	9	43	31	53	46	51	47	-4
59CR	14	37	6	39	9	31	10	30	8	22	39	15	56	31	48	34	-14
61CR	47	62	65	86	14	39	7	40	24	42	23	6	20	5	28	6	-22
64CR	11	34	14	51	19	30	10	26	5	7	41	22	44	30	48	34	-14
66CR	10	57	10	78	17	58	10	56	5	56	40	7	51	7	47	15	-32
67CR	35	50	26	56	8	39	5	54	10	37	43	21	34	19	42	25	-17
68CR	14	56	NA	NA	9	80	5	74	8	19	42	24	50	12	48	18	-30
74CR	7	61	2	19	5	55	1	61	4	26	35	5	39	11	45	12	-33
75CR	15	44	13	34	7	59	5	63	9	54	50	9	40	11	50	13	-37
76CR	21	37	4	67	9	48	4	46	NA	NA	32	15	42	7	45	27	-18
77CR	26	39	12	51	13	44	11	41	9	26	48	19	53	23	46	13	-33
81CR	11	53	7	45	14	49	8	33	5	53	38	8	55	11	55	18	-37
85CR	5	67	9	46	6	69	6	63	12	17	42	33	37	5	48	11	-37
86CR	13	30	16	61	12	55	13	40	7	16	45	19	51	17	43	24	-19
87CR	12	23	13	28	8	30	8	28	10	21	50	26	48	32	53	41	-12
92CR	25	71	17	58	19	67	7	68	16	43	40	19	36	9	45	12	-33
96CR	25	65	9	58	14	51	10	41	9	30	35	13	45	10	48	12	-36

ID	ZNF.N	ZNF.T	KCNQ5.N	KCNQ5.T	SPOCK5.N	SPOCK5.T	SPOCK7.N	SPOCK7.T	LOC.N	LOC.T	Chr1.N	Chr1.T	Chr8.N	Chr8.T	Chr14.N	Chr14.T	Chr14.D
97CR	2	49	1	53	3	53	3	53	7	47	57	10	47	17	NA	NA	NA
98CR	12	59	11	62	20	67	15	66	12	57	29	8	43	8	52	26	-26
99CR	4	28	8	43	4	19	6	25	7	34	41	20	NA	NA	22	33	11
100CR	1	8	2	16	11	44	5	67	4	6	49	24	38	7	37	26	-11
101CR	22	42	16	44	15	13	15	6	13	34	40	13	46	21	36	19	-17
102CR	15	55	6	57	9	52	7	54	8	52	32	11	23	10	45	24	-21
103CR	11	63	8	53	9	33	7	25	5	40	46	23	44	13	47	26	-21
106CR	10	49	9	53	10	53	6	36	11	33	38	11	42	10	50	15	-35
110CR	21	77	7	70	11	73	6	72	10	74	41	9	38	7	45	10	-35
111CR	15	43	5	45	15	20	6	7	8	3	40	38	37	9	57	26	-31
113CR	15	40	8	60	50	42	7	37	10	45	47	12	44	13	41	21	-20
114CR	7	56	7	33	11	29	7	28	8	31	41	9	32	2	42	20	-22
116CR	6	45	7	45	10	47	6	36	5	34	50	14	38	22	25	30	5
119CR	16	50	11	54	25	49	21	48	9	34	25	11	48	22	41	20	-21
120CR	28	69	23	76	21	77	13	74	14	20	49	8	47	4	45	11	-34

122CR	20	66	5	72	13	64	11	63	9	57	47	21	49	7	47	32	-15
123CR	10	45	12	50	12	7	9	4	10	14	38	38	44	5	51	11	-40
124CR	15	46	13	40	18	43	15	10	15	8	49	11	48	11	48	16	-32
126CR	9	39	4	55	12	65	4	59	8	50	36	12	44	19	38	17	-21
127CR	33	78	10	73	14	71	10	71	12	66	24	28	35	17	39	46	7
128CR	25	28	15	28	32	31	15	18	14	18	32	18	34	21	38	28	-10
129CR	41	50	36	40	37	52	22	51	28	24	40	29	21	23	41	36	-5
132CR	2	4	5	11	6	10	9	9	7	9	48	11	33	26	26	20	-6

Appendix 4: Clinicopathological features of historical cohort

ID	Sex	Age	Site	Side	Dukes	Differentiation	T stage	N stage	Metastasis	EMVI	Death within 5 years
3CR	M	49	Recto-sigmoid	L	B	Mod	3	0	N	N	Y
4CR	M	75	Caecum	R	D	Well/mod	3	1	Y	N	Y
6CR	M	77	Caecum	R	B	Well/mod	3	0	N	N	N
8CR	M	49	Distal sigmoid	L	A	Well/mod	1	0	N	N	N
11CR	F	71	Distal sigmoid	L	B	Well/mod	3	0	N	N	N
13CR	F	71	Anorectal junction	L	B	Mod	3	1	N	N	Y
15CR	M	86	Caecum	R	A	Mod	2	0	N	N	N
20CR	F	54	Descending colon	L	C	Well/mod	3	1	N	N	N
24CR	M	83	Sigmoid	L	B	Well/mod	4	0	N	N	N
25CR	M	69	Rectosigmoid junction	L	B	Mod	3	0	N	N	Y
28CR	F	77	Caecum	R	B	Well/mod	3	0	N	Y	N
30CR	F	71	Distal sigmoid	L	B	Well/mod	3	0	N	N	N
31CR	F	83	Distal sigmoid	L	C	Mod	3	1	N	N	Y
33CR	M	88	Low rectum	L	C	Well/mod	2	1	N	N	N
35CR	F	72	Distal sigmoid	L	A	Mod	2	0	N	N	N
37CR	F	79	Caecum	R	B	Well/mod	4	0	N	N	N

38CR	F	85	Proximal transverse	R	C	Well/mod	3	1	N	N	N
39CR	F	61	Mid sigmoid	L	B	Well/mod	3	0	N	N	N
40CR	M	85	Caecum	R	B	Well/mod	3	0	N	N	Y
41CR	F	71	Rectosigmoid junction	L	C	Well/mod	3	1	N	N	N
42CR	M	84	Caecum	R	B	Well/mod	3	0	N	N	N
43CR	F	76	Distal transverse	L	B	Mod	3	0	N	N	N
44CR	F	82	Caecum	R	B	Well/mod	3	0	N	N	N
45CR	M	82	caecum	R	C	Well/mod	3	1	N	N	Y
47CR	F	72	Caecum	R	B	Well/mod	4	0	N	Y	N
51CR	M	75	Low rectum	L	B	Well/mod	3	0	N	N	N
52CR	F	84	Caecum	R	B	Well/mod	3	0	N	N	N
56CR	F	57	Ascending colon	R	B	Well/mod	4	0	N	N	N
57CR	F	60	Mid sigmoid	L	B	Well/mod	4	0	N	Y	N
59CR	M	87	Rectal	L	B	Well/mod	3	0	N	N	N
61CR	M	89	Caecum	R	B	Well/mod	3	0	N	N	Y
64CR	M	83	High rectum	L	B	Well/mod	3	0	N	N	N
66CR	M	66	Caecum	R	C	Well/mod	3	1	N	N	N
67CR	M	85	Caecum	R	B	Poor	3	0	N	N	Y
68CR	M	71	Rectosigmoid junction	L	B	Well/mod	3	0	N	Y	N
74CR	F	80	Proximal sigmoid	L	B	Well/mod	3	0	N	N	N
75CR	M	66	Transverse	R	C	Well/mod	3	2	N	Y	N
76CR	F	81	Caecum	R	B	Well/mod	3	0	N	N	N

77CR	F	75	Hepatic flexure	R	B	Well/mod	3	0	N	N	N
81CR	F	73	Mid sigmoid	L	C	Well/mod	3	1	N	N	N
85CR	M	61	Lower sigmoid	L	B	Well/mod	3	0	N	N	N
86CR	M	73	Descending	L	C	Well/mod	4	1	N	Y	N
87CR	F	77	Distal ascending colon	R	B	Well/mod	3	0	N	N	N
92CR	M	67	High Rectal	L	A	Well/mod	2	0	N	N	N
96CR	F	87	Caecum	R	B	Well/mod	3	0	N	N	Y
97CR	M	69	Caecum	R	B	Well/mod	3	0	N	N	N
98CR	F	71	Rectosigmoid	L	B	Well/mod	4	0	N	N	N
99CR	M	77	Ascending colon	R	B	Well/mod	3	0	N	N	N
100CR	M	68	Caecum	R	B	Well/mod	3	0	N	N	N
101CR	F	65	Descending colon	L	C	Poor	4	2	N	N	Y
102CR	F	70	High rectal	L	B	Well/mod	3	0	N	N	N
103CR	M	51	Sigmoid	L	A	Mod	2	0	N	N	N
106CR	M	59	Distal sigmoid	L	B	Well/mod	3	0	N	N	N
110CR	M	82	Proximal transverse	R	C	Poor	4	1	N	Y	Y
111CR	M	76	Upper rectum	L	B	Well/mod	3	0	N	N	N
113CR	M	81	Ascending colon	R	B	Mod	3	0	N	N	N

114CR	F	75	Distal sigmoid	L	B	Well/mod	3	0	N	N	N
116CR	F	86	Hepatic flexure	R	C	Well/mod	3	1	N	N	N
119CR	F	59	Ascending colon	R	D	Mod	4	2	Y	Y	Y
120CR	F	80	Ascending colon	R	C	Well/mod	3	1	N	N	Y
122CR	M	71	Ascending colon	R	B	Mod	2	0	N	N	N
123CR	M	64	Sigmoid	L	B	Well/mod	3	0	N	N	N
124CR	M	64	Distal sigmoid	L	B	Well/mod	4	0	N	Y	N
126CR	M	65	Caecal	R	B	Well/mod	3	0	N	N	N
127CR	F	75	Transverse	R	B	Poor	3	0	N	N	N
128CR	F	62	Low rectum	L	B	Well/mod	2	0	N	Y	N
129CR	F	63	Hepatic flexure	R	C	Well/mod	3	1	N	N	N
132CR	M	62	Rectum	L	C	Mod	3	1	N	N	N
	Supplementary Table 4: Clinicopathological features of historical cohort. M, male; F, female; L, left; R, right; N, no; Y, yes.										

Bibliography

1. Maleszewska M, Kaminska B. Is glioblastoma an epigenetic malignancy? Cancers. 2013;5(3):1120-39.
2. Felsenfeld G. A Brief History of Epigenetics. Cold Spring Harbor Perspectives in Biology. 2014;6(1): a018200.
3. Van Loosdregt J, Coffer PJ. Post-translational modification networks regulating FOXP3 function. Trends Immunol. 2014;35(8):368-78.
4. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, Yurovsky A, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLoS genetics. 2015;11(1):e1004958.
5. Byun HM, Nordio F, Coull BA, Tarantini L, Hou L, Bonzini M, et al. Temporal stability of epigenetic markers: sequence characteristics and predictors of short-term DNA methylation variations. PloS one. 2012;7(6):e39220.
6. Chen Y, Breeze CE, Zhen S, Beck S, Teschendorff AE. Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. Epigenetics & chromatin. 2016;9:10.
7. Huang C, Wen B. "Identification Card": Sites on Histone Modification of Cancer Cell. Chinese medical sciences journal = Chung-kuo i hsueh k'o hsueh tsa chih. 2015;30(4):203-9.
8. Chen T, Dent SYR. Chromatin modifiers: regulators of cellular differentiation. Nature reviews Genetics. 2014;15(2):93-106.
9. Castel SE, Martienssen RA. RNA interference (RNAi) in the Nucleus: roles for small RNA in transcription, epigenetics and beyond. Nature reviews Genetics. 2013;14(2):100-12.
10. Buschbeck M, Hake SB. Variants of core histones and their roles in cell fate decisions, development and cancer. Nature reviews Molecular cell biology. 2017;18(5):299-314.
11. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799-816.
12. Boyes J, Bird A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell. 1991;64(6):1123-34.
13. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nature reviews Genetics. 2011;13(1):7-13.
14. Koppe MJ, Boerman OC, Oyen WJG, Bleichrodt RP. Peritoneal Carcinomatosis of Colorectal Origin: Incidence and Current Treatment Strategies. Annals of surgery. 2006;243(2):212-22.
15. Ali AA, Marcus JN, Harvey JP, Roll R, Hodgson CP, Wildrick DM, et al. RB1 protein in normal and malignant human colorectal tissue and colon cancer cell lines. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 1993;7(10):931-7.
16. Iacopetta B. TP53 mutation in colorectal cancer. Human mutation. 2003;21(3):271-6.
17. Fodde R. The APC gene in colorectal cancer. European journal of cancer (Oxford, England : 1990). 2002;38(7):867-71.
18. Araten DJ, Golde DW, Zhang RH, Thaler HT, Gargiulo L, Notaro R, et al. A quantitative measurement of the human somatic mutation rate. Cancer research. 2005;65(18):8111-7.
19. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science (New York, NY). 2015;347(6217):78-81.

20. Tarabichi M. A research note regarding "Variation in cancer risk among tissues can be explained by the number of stem cell divisions". F1000Research. 2016;5:2044.
21. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proceedings of the National Academy of Sciences of the United States of America. 1971;68(4):820-3.
22. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
23. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
24. Vasovcak P, Krepelova A, Menigatti M, Puchmajerova A, Skapa P, Augustinakova A, et al. Unique mutational profile associated with a loss of TDG expression in the rectal cancer of a patient with a constitutional PMS2 deficiency. DNA repair. 2012;11(7):616-23.
25. Standring SB, N. Collins, P. Crossman, AR. Gatzoulis, MA. Healy, JC. Johnson, D. Mahadevan, V. Newell RLM. Wigley, C. Gray's Anatomy: The Anatomical Basis of Clinical Practice. 40 ed. London: Elsevier; 2008. p 1137-62.
26. Kenig J, Richter P. Definition of the rectum and level of the peritoneal reflection - still a matter of debate? Wideochirurgia i inne techniki maloinwazyjne = Videosurgery and other miniinvasive techniques. 2013;8(3):183-6.
27. Ireland AoCoGBa. Colorectal Cancer Management Guidelines 3ed2007.
28. Montgomery RK, Mulberg AE, Grand RJ. Development of the human gastrointestinal tract: twenty years of progress. Gastroenterology. 1999;116(3):702-31.
29. Coffey JC, O'Leary DP. The mesentery: structure, function, and role in disease. The lancet Gastroenterology & hepatology. 2016;1(3):238-47.
30. Heald RJ, Ryall RD. Recurrence and survival after total mesorectal excision for rectal cancer. Lancet (London, England). 1986;1(8496):1479-82.
31. Mukkai Krishnamurty D, Wise PE. Importance of surgical margins in rectal cancer. Journal of surgical oncology. 2016;113(3):323-32.
32. Salerno G, Daniels IR, Moran BJ, Wotherspoon A, Brown G. Clarifying margins in the multidisciplinary management of rectal cancer: the MERCURY experience. Clinical radiology. 2006;61(11):916-23.
33. Shamsuddin AM, Phelps PC, Trump BF. Human large intestinal epithelium: light microscopy, histochemistry, and ultrastructure. Human pathology. 1982;13(9):790-803.
34. Dudeja PK, Rao DD, Syed I, Joshi V, Dahdal RY, Gardner C, et al. Intestinal distribution of human Na+/H+ exchanger isoforms NHE-1, NHE-2, and NHE-3 mRNA. The American journal of physiology. 1996;271(3 Pt 1):G483-93.
35. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015;136(5):E359-86.
36. UK CR. Bowel cancer incidence statistics 2017 [
37. Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C, Kohler B, et al. Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. International journal of cancer. 2015;137(9):2060-71.
38. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. Nature reviews Disease primers. 2015;1:15065.
39. Mima K, Sukawa Y, Nishihara R, Qian ZR, Yamauchi M, Inamura K, et al. Fusobacterium nucleatum and T Cells in Colorectal Carcinoma. JAMA oncology. 2015;1(5):653-61.
40. Chubak J, Kamineni A, Buist DSM, Anderson ML, Whitlock EP. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Aspirin Use for the Prevention of Colorectal Cancer: An Updated Systematic Evidence Review for

the US Preventive Services Task Force. Rockville (MD): Agency for Healthcare Research and Quality (US); 2015.
41. Liao X, Lochhead P, Nishihara R, Morikawa T, Kuchiba A, Yamauchi M, et al. Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. The New England journal of medicine. 2012;367(17):1596-606.
42. Makar GA, Holmes JH, Yang YX. Angiotensin-Converting Enzyme Inhibitor Therapy and Colorectal Cancer Risk. Journal of the National Cancer Institute. 2014; 106(2): djt374.
43. Kloor M, Staffa L, Ahadova A, von Knebel Doeberitz M. Clinical significance of microsatellite instability in colorectal cancer. Langenbeck's archives of surgery. 2014;399(1):23-31.
44. Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology. 2010;138(6):2059-72.
45. Scholefield JH. ABC of colorectal cancer: screening. BMJ (Clinical research ed). 2000;321(7267):1004-6.
46. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science (New York, NY). 2006;314(5797):268-74.
47. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23(3):609-18.
48. FC DAS, Wernhoff P, Dominguez-Barrera C, Dominguez-Valentin M. Update on Hereditary Colorectal Cancer. Anticancer research. 2016;36(9):4399-405.
49. Lowery JT, Ahnen DJ, Schroy PC, 3rd, Hampel H, Baxter N, Boland CR, et al. Understanding the contribution of family history to colorectal cancer risk and its clinical implications: A state-of-the-science review. Cancer. 2016;122(17):2633-45.
50. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-analyses of Colorectal Cancer Risk Factors. Cancer causes & control : CCC. 2013;24(6):1207-22.
51. Li J, Huang KL, Zhang T, Li H, Zhao J, Wang H. Pan-cancer methylation and expression profiling of adenocarcinomas revealed epigenetic silencing in the WNT signaling pathway. Neoplasma. 2016;63(2):208-14.
52. Ponder BA. Cancer genetics. Nature. 2001;411(6835):336-41.
53. Tanikawa C, Kamatani Y, Takahashi A, Momozawa Y, Leveque K, Nagayama S, et al. GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12. Carcinogenesis. 2018;39(5):652-60.
54. Cui R, Okada Y, Jang SG, Ku JL, Park JG, Kamatani Y, et al. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. Gut. 2011;60(6):799-805.
55. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nature genetics. 2010;42(11):973-7.
56. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nature genetics. 2008;40(12):1426-35.
57. Peters U, Bien S, Zubair N. Genetic Architecture of Colorectal Cancer. Gut. 2015;64(10):1623-36.
58. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89-92.

59. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethylation of DNA from benign and malignant human colon neoplasms. Science (New York, NY). 1985;228(4696):187-90.
60. Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, et al. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut. 2014;63(4):635-46.
61. Benard A, van de Velde CJ, Lessard L, Putter H, Takeshima L, Kuppen PJ, et al. Epigenetic status of LINE-1 predicts clinical outcome in early-stage rectal cancer. British journal of cancer. 2013;109(12):3073-83.
62. Nosho K, Baba Y, Tanaka N, Shima K, Hayashi M, Meyerhardt JA, et al. Tumour-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. The Journal of pathology. 2010;222(4):350-66.
63. Holm TM, Jackson-Grusby L, Brambrink T, Yamada Y, Rideout WM, 3rd, Jaenisch R. Global loss of imprinting leads to widespread tumorigenesis in adult mice. Cancer cell. 2005;8(4):275-85.
64. Slattery ML, Wolff RK, Herrick JS, Curtin K, Caan BJ, Samowitz W. Alcohol consumption and rectal tumor mutations and epigenetic changes. Diseases of the colon and rectum. 2010;53(8):1182-9.
65. Paun BC, Kukuruga D, Jin Z, Mori Y, Cheng Y, Duncan M, et al. Relation between normal rectal methylation, smoking status, and the presence or absence of colorectal adenomas. Cancer. 2010;116(19):4495-501.
66. Wallace K, Grau MV, Levine AJ, Shen L, Hamdan R, Chen X, et al. Association between folate levels and CpG Island hypermethylation in normal colorectal mucosa. Cancer prevention research (Philadelphia, Pa). 2010;3(12):1552-64.
67. Morikawa T, Kuchiba A, Yamauchi M, Meyerhardt JA, Shima K, Nosho K, et al. Association of CTNNB1 (beta-catenin) alterations, body mass index, and physical activity with survival in patients with colorectal cancer. Jama. 2011;305(16):1685-94.
68. Hanyuda A, Kim SA, Martinez-Fernandez A, Qian ZR, Yamauchi M, Nishihara R, et al. Survival Benefit of Exercise Differs by Tumor IRS1 Expression Status in Colorectal Cancer. Annals of surgical oncology. 2016;23(3):908-17.
69. Verma M, Kumar V. Epigenetic Biomarkers in Colorectal Cancer. Molecular diagnosis & therapy. 2017;21(2):153-65.
70. Tanikawa C, Furukawa Y, Yoshida N, Arakawa H, Nakamura Y, Matsuda K. XEDAR as a putative colorectal tumor suppressor that mediates p53-regulated anoikis pathway. Oncogene. 2009;28(34):3081-92.
71. Ying J, Poon FF, Yu J, Geng H, Wong AH, Qiu GH, et al. DLEC1 is a functional 3p22.3 tumour suppressor silenced by promoter CpG methylation in colon and gastric cancers. British journal of cancer. 2009;100(4):663-9.
72. Dai CY, Furth EE, Mick R, Koh J, Takayama T, Niitsu Y, et al. p16(INK4a) expression begins early in human colon neoplasia and correlates inversely with markers of cell proliferation. Gastroenterology. 2000;119(4):929-42.
73. Kamory E, Olasz J, Csuka O. Somatic APC inactivation mechanisms in sporadic colorectal cancer cases in Hungary. Pathology oncology research : POR. 2008;14(1):51-6.
74. Curtin K, Ulrich CM, Samowitz WS, Wolff RK, Duggan DJ, Makar KW, et al. Candidate pathway polymorphisms in one-carbon metabolism and risk of rectal tumor mutations. International journal of molecular epidemiology and genetics. 2011;2(1):1-8.
75. Molinari C, Casadio V, Foca F, Zingaretti C, Giannini M, Avanzolini A, et al. Gene methylation in rectal cancer: predictive marker of response to chemoradiotherapy? Journal of cellular physiology. 2013;228(12):2343-9.

76. Leong KJ, Wei W, Tannahill LA, Caldwell GM, Jones CE, Morton DG, et al. Methylation profiling of rectal cancer identifies novel markers of early-stage disease. The British journal of surgery. 2011;98(5):724-34.
77. Gaedcke J, Leha A, Claus R, Weichenhan D, Jung K, Kitz J, et al. Identification of a DNA methylation signature to predict disease-free survival in locally advanced rectal cancer. Oncotarget. 2014;5(18):8123-35.
78. Naumov VA, Generozov EV, Zaharjevskaya NB, Matushkina DS, Larin AK, Chernyshov SV, et al. Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. Epigenetics. 2013;8(9):921-34.
79. Ha YJ, Kim CW, Roh SA, Cho DH, Park JL, Kim SY, et al. Epigenetic regulation of KLHL34 predictive of pathologic response to preoperative chemoradiation therapy in rectal cancer patients. International journal of radiation oncology, biology, physics. 2015;91(3):650-8.
80. Lin PC, Lin JK, Lin CH, Lin HH, Yang SH, Jiang JK, et al. Clinical Relevance of Plasma DNA Methylation in Colorectal Cancer Patients Identified by Using a Genome-Wide High-Resolution Array. Annals of surgical oncology. 2015;22 Suppl 3:S1419-27.
81. Vymetalkova V, Vodicka P, Pardini B, Rosa F, Levy M, Schneiderova M, et al. Epigenome-wide analysis of DNA methylation reveals a rectal cancer-specific epigenomic signature. Epigenomics. 2016;8(9):1193-207.
82. Wei J, Li G, Zhang J, Zhou Y, Dang S, Chen H, et al. Integrated analysis of genome-wide DNA methylation and gene expression profiles identifies potential novel biomarkers of rectal cancer. Oncotarget. 2016;7(38):62547-58.
83. Jia M, Gao X, Zhang Y, Hoffmeister M, Brenner H. Different definitions of CpG island methylator phenotype and outcomes of colorectal cancer: a systematic review. Clinical epigenetics. 2016;8.
84. Bae JM, Kim JH, Cho NY, Kim TY, Kang GH. Prognostic implication of the CpG island methylator phenotype in colorectal cancers depends on tumour location. British journal of cancer. 2013;109(4):1004-12.
85. Jo P, Jung K, Grade M, Conradi LC, Wolff HA, Kitz J, et al. CpG island methylator phenotype infers a poor disease-free survival in locally advanced rectal cancer. Surgery. 2012;151(4):564-70.
86. Williamson JS, Jones HG, Williams N, Griffiths AP, Jenkins G, Beynon J, et al. Extramural vascular invasion and response to neoadjuvant chemoradiotherapy in rectal cancer: Influence of the CpG island methylator phenotype. World journal of gastrointestinal oncology. 2017;9(5):209-17.
87. Nishihara R, Wu K, Lochhead P, Morikawa T, Liao X, Qian ZR, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. The New England journal of medicine. 2013;369(12):1095-105.
88. Hinoue T, Weisenberger DJ, Pan F, Campan M, Kim M, Young J, et al. Analysis of the association between CIMP and BRAF in colorectal cancer by DNA methylation profiling. PloS one. 2009;4(12):e8357.
89. Yap AS, Brieher WM, Gumbiner BM. Molecular and functional analysis of cadherin-based adherens junctions. Annual review of cell and developmental biology. 1997;13:119-46.
90. Greenburg G, Hay ED. Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. The Journal of cell biology. 1982;95(1):333-9.
91. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nature reviews Cancer. 2009;9(4):265-73.

92. Wu CY, Tsai YP, Wu MZ, Teng SC, Wu KJ. Epigenetic reprogramming and post-transcriptional regulation during the epithelial-mesenchymal transition. Trends in genetics : TIG. 2012;28(9):454-63.
93. Wang JQ, Wu KJ. Epigenetic regulation of epithelial-mesenchymal transition by hypoxia in cancer: targets and therapy. Current pharmaceutical design. 2015;21(10):1272-8.
94. Leong KJ, Beggs A, James J, Morton DG, Matthews GM, Bach SP. Biomarker-based treatment selection in early-stage rectal cancer to promote organ preservation. The British journal of surgery. 2014;101(10):1299-309.
95. Benard A, Zeestraten EC, Goossens-Beumer IJ, Putter H, van de Velde CJ, Hoon DS, et al. DNA methylation of apoptosis genes in rectal cancer predicts patient survival and tumor recurrence. Apoptosis : an international journal on programmed cell death. 2014;19(11):1581-93.
96. de Maat MF, van de Velde CJ, van der Werff MP, Putter H, Umetani N, Klein-Kranenbarg EM, et al. Quantitative analysis of methylation of genomic loci in early-stage rectal cancer predicts distant recurrence. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(14):2327-35.
97. de Maat MF, van de Velde CJ, Benard A, Putter H, Morreau H, van Krieken JH, et al. Identification of a quantitative MINT locus methylation profile predicting local regional recurrence of rectal cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010;16(10):2811-8.
98. Kohonen-Corish MR, Tseung J, Chan C, Currey N, Dent OF, Clarke S, et al. KRAS mutations and CDKN2A promoter methylation show an interactive adverse effect on survival and predict recurrence of rectal cancer. International journal of cancer. 2014;134(12):2820-8.
99. Vymetalkova VP, Slyskova J, Korenkova V, Bielik L, Langerova L, Prochazka P, et al. Molecular characteristics of mismatch repair genes in sporadic colorectal tumors in Czech patients. BMC medical genetics. 2014;15:17.
100. Exner R, Pulverer W, Diem M, Spaller L, Woltering L, Schreiber M, et al. Potential of DNA methylation in rectal cancer as diagnostic and prognostic biomarkers. British journal of cancer. 2015;113(7):1035-45.
101. Laskar RS, Ghosh SK, Talukdar FR. Rectal cancer profiling identifies distinct subtypes in India based on age at onset, genetic, epigenetic and clinicopathological characteristics. Molecular carcinogenesis. 2015;54(12):1786-95.
102. Liu JB, Zhou QB, Xu JZ, Wang GX, Yuan WT. Influence of colorectal cancer tumor suppressor gene CHD5 methylation on its clinical and pathological characteristics. Journal of biological regulators and homeostatic agents. 2015;29(4):889-93.
103. Hua Y, Ma X, Liu X, Yuan X, Qin H, Zhang X. Abnormal expression of mRNA, microRNA alteration and aberrant DNA methylation patterns in rectal adenocarcinoma. PloS one. 2017;12(3):e0174461.
104. Yokoi K, Yamashita K, Ishii S, Tanaka T, Nishizawa N, Tsutsui A, et al. Comprehensive molecular exploration identified promoter DNA methylation of the CRBP1 gene as a determinant of radiation sensitivity in rectal cancer. British journal of cancer. 2017;116(8):1046-56.
105. Rodel C, Hofheinz R, Liersch T. Rectal cancer: state of the art in 2012. Current opinion in oncology. 2012;24(4):441-7.
106. McCarthy K, Pearson K, Fulton R, Hewitt J. Pre-operative chemoradiation for non-metastatic locally advanced rectal cancer. The Cochrane database of systematic reviews. 2012;12:Cd008368.
107. Doubeni CA, Major JM, Laiyemo AO, Schootman M, Zauber AG, Hollenbeck AR, et al. Contribution of Behavioral Risk Factors and Obesity to Socioeconomic Differences in Colorectal Cancer Incidence. Journal of the National Cancer Institute. 2012;104(18):1353-62.

108. Matsuda T, Yamashita K, Hasegawa H, Oshikiri T, Hosono M, Higashino N, et al. Recent updates in the surgical treatment of colorectal cancer. Annals of Gastroenterological Surgery. 2018;2(2):129-36.
109. Slattery ML, Curtin K, Wolff RK, Boucher KM, Sweeney C, Edwards S, et al. A comparison of colon and rectal somatic DNA alterations. Diseases of the colon and rectum. 2009;52(7):1304-11.
110. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.
111. Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. Annals of internal medicine. 1990;113(10):779-88.
112. Yamauchi M, Morikawa T, Kuchiba A, Imamura Y, Qian ZR, Nishihara R, et al. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. Gut. 2012;61(6):847-54.
113. Ogino S, Lochhead P, Giovannucci E, Meyerhardt JA, Fuchs CS, Chan AT. Discovery of colorectal cancer PIK3CA mutation as potential predictive biomarker: power and promise of molecular pathological epidemiology. Oncogene. 2014;33(23):2949-55.
114. Lee MS, Menter DG, Kopetz S. Right Versus Left Colon Cancer Biology: Integrating the Consensus Molecular Subtypes. Journal of the National Comprehensive Cancer Network : JNCCN. 2017;15(3):411-9.
115. Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017;66(4):633-43.
116. Qin Y, Wade PA. Crosstalk between the microbiome and epigenome: messages from bugs. Journal of biochemistry. 2018;163(2):105-12.
117. Thomas LA, Veysey MJ, French G, Hylemon PB, Murphy GM, Dowling RH. Bile acid metabolism by fresh human colonic contents: a comparison of caecal versus faecal samples. Gut. 2001;49(6):835-42.
118. Saracut C, Molnar C, Russu C, Todoran N, Vlase L, Turdean S, et al. Secondary bile acids effects in colon pathology. Experimental mice study. Acta cirurgica brasileira. 2015;30(9):624-31.
119. Glebov OK, Rodriguez LM, Nakahara K, Jenkins J, Cliatt J, Humbyrd CJ, et al. Distinguishing right from left colon by the pattern of gene expression. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2003;12(8):755-62.
120. Gonsalves WI, Mahoney MR, Sargent DJ, Nelson GD, Alberts SR, Sinicrope FA, et al. Patient and tumor characteristics and BRAF and KRAS mutations in colon cancer, NCCTG/Alliance N0147. Journal of the National Cancer Institute. 2014;106(7).
121. Weiss JM, Pfau PR, O'Connor ES, King J, LoConte N, Kennedy G, et al. Mortality by stage for right- versus left-sided colon cancer: analysis of surveillance, epidemiology, and end results--Medicare data. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(33):4401-9.
122. Zhang RX, Ma WJ, Gu YT, Zhang TQ, Huang ZM, Lu ZH, et al. Primary tumor location as a predictor of the benefit of palliative resection for colorectal cancer with unresectable metastasis. World journal of surgical oncology. 2017;15(1):138.
123. Cao DD, Xu HL, Xu XM, Ge W. The impact of primary tumor location on efficacy of cetuximab in metastatic colorectal cancer patients with different Kras status: a systematic review and meta-analysis. Oncotarget. 2017;8(32):53631-41.

124. Holch JW, Ricard I, Stintzing S, Modest DP, Heinemann V. The relevance of primary tumour location in patients with metastatic colorectal cancer: A meta-analysis of first-line clinical trials. European journal of cancer (Oxford, England : 1990). 2017;70:87-98.
125. Ait Ouakrim D, Pizot C, Boniol M, Malvezzi M, Boniol M, Negri E, et al. Trends in colorectal cancer mortality in Europe: retrospective analysis of the WHO mortality database. BMJ (Clinical research ed). 2015;351:h4970.
126. Kubisch CH, Crispin A, Mansmann U, Goke B, Kolligs FT. Screening for Colorectal Cancer Is Associated With Lower Disease Stage: A Population-Based Study. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2016;14(11):1612-8.e3.
127. Facciorusso A, Ferrusquia J, Muscatiello N. Lead time bias in estimating survival outcomes. Gut. 2016;65(3):538-9.
128. Binefa G, Rodriguez-Moranta F, Teule A, Medina-Hayas M. Colorectal cancer: from prevention to personalized medicine. World journal of gastroenterology. 2014;20(22):6786-808.
129. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. 2015;64(10):1637-49.
130. Shima K, Morikawa T, Baba Y, Nosho K, Suzuki M, Yamauchi M, et al. MGMT promoter methylation, loss of expression and prognosis in 855 colorectal cancers. Cancer causes & control : CCC. 2011;22(2):301-9.
131. Grutzmann R, Molnar B, Pilarsky C, Habermann JK, Schlag PM, Saeger HD, et al. Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. PloS one. 2008;3(11):e3759.
132. Heiss JA, Brenner H. Epigenome-wide discovery and evaluation of leukocyte DNA methylation markers for the detection of colorectal cancer in a screening setting. Clinical epigenetics. 2017;9:24.
133. Bai M, Yu NZ, Long F, Feng C, Wang XJ. Effects of CDKN2A (p16INK4A/p14ARF) Over-Expression on Proliferation and Migration of Human Melanoma A375 Cells. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology. 2016;40(6):1367-76.
134. Pellise M, Castells A, Gines A, Agrelo R, Sole M, Castellvi-Bel S, et al. Detection of lymph node micrometastases by gene promoter hypermethylation in samples obtained by endosonography- guided fine-needle aspiration biopsy. Clinical cancer research : an official journal of the American Association for Cancer Research. 2004;10(13):4444-9.
135. Kaminski MF, Regula J, Kraszewska E, Polkowski M, Wojciechowska U, Didkowska J, et al. Quality indicators for colonoscopy and the risk of interval cancer. The New England journal of medicine. 2010;362(19):1795-803.
136. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. Journal of the National Cancer Institute. 2004;96(19):1420-5.
137. Network. NCC. NCCN Clinical Practice Guidelines in Oncology: Rectal Cancer. Version 2.2016 ed2016.
138. NCIN. Colorectal Cancer Survival by Stage - NCIN Data Briefing 2010 [Available from: http://www.ncin.org.uk/publications/data_briefings/colorectal_cancer_survival_by_stage.
139. Sinicrope FA, Shi Q, Smyrk TC, Thibodeau SN, Dienstmann R, Guinney J, et al. Molecular markers identify subtypes of stage III colon cancer associated with patient outcomes. Gastroenterology. 2015;148(1):88-99.

140. Fakhri B, Lim KH. Molecular landscape and sub-classification of gastrointestinal cancers: a review of literature. Journal of Gastrointestinal Oncology. 2017;8(3):379-86.
141. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The Consensus Molecular Subtypes of Colorectal Cancer. Nature medicine. 2015;21(11):1350-6.
142. Sveen A, Bruun J, Eide PW, Eilertsen IA, Ramirez L, Murumagi A, et al. Colorectal Cancer Consensus Molecular Subtypes Translated to Preclinical Models Uncover Potentially Targetable Cancer Cell Dependencies. Clinical cancer research : an official journal of the American Association for Cancer Research. 2018;24(4):794-806.
143. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics (Oxford, England). 2012;28(22):2986-8.
144. Desjardins P, Conklin D. NanoDrop microvolume quantitation of nucleic acids. Journal of visualized experiments : JoVE. 2010(45).
145. Qiagen. A New Standard in Epigenetic and Genetic Analysis! In: Qiagen, editor. 2010.
146. Corp ZR. Instruction manual: EZ-96 DNA Methylation-GoldTM Kit. Available from: http://www.zymoresearch.com/downloads/dl/file/id/59/d5007i.pdf.
147. CliniSciences. Conventional PCR 2017 [Available from: https://www.clinisciences.com/en/buy/cat-conventional-pcr-3473.html.
148. Cruz UoCS. UCSC Genome Browser 2009 [Available from: https://genome-euro.ucsc.edu/cgi-bin/hgGateway?hgsid=602792263_kALNBQw8IlFpyAl3L56AgEWs3fC2&redirect=manual&source=genome.ucsc.edu.
149. Sanderson BA, Araki N, Lilley JL, Guerrero G, Lewis LK. Modification of gel architecture and TBE/TAE buffer composition to minimize heating during agarose gel electrophoresis. Analytical biochemistry. 2014;454:44-52.
150. Tost J, Gut IG. DNA methylation analysis by pyrosequencing. Nature protocols. 2007;2(9):2265-75.
151. Karimi M, Johansson S, Ekstrom TJ. Using LUMA: a Luminometric-based assay for global DNA-methylation. Epigenetics. 2006;1(1):45-8.
152. Wei H, Therrien C, Blanchard A, Guan S, Zhu Z. The Fidelity Index provides a systematic quantitation of star activity of DNA restriction endonucleases. Nucleic acids research. 2008;36(9):e50.
153. Lisanti S, Omar WAW, Tomaszewski B, De Prins S, Jacobs G, Koppen G, et al. Comparison of Methods for Quantification of Global DNA Methylation in Human Cells and Tissues. PloS one. 2013;8(11).
154. Duman EA, Kriaucionis S, Dunn JJ, Hatchwell E. A simple modification to the luminometric methylation assay to control for the effects of DNA fragmentation. BioTechniques. 2015;58(5):262-4.
155. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, et al. Intra-individual change in DNA methylation over time with familial clustering. Jama. 2008;299(24):2877-83.
156. Knothe C, Shiratori H, Resch E, Ultsch A, Geisslinger G, Doehring A, et al. Disagreement between two common biomarkers of global DNA methylation. Clinical epigenetics. 2016;8:60.
157. Sant KE, Nahar MS, Dolinoy DC. DNA Methylation Screening and Analysis. Methods in molecular biology (Clifton, NJ). 2012;889:385-406.
158. Canene-Adams K. Preparation of formalin-fixed paraffin-embedded tissue for immunohistochemistry. Methods in enzymology. 2013;533:225-33.

159. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clinical chemistry. 2015;61(1):64-71.
160. Yan L, Sun K, Liu Y, Liang J, Cai K, Gui J. MiR-129-5p influences the progression of gastric cancer cells through interacting with SPOCK1. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2017;39(6):1010428317706916.
161. Shu YJ, Weng H, Ye YY, Hu YP, Bao RF, Cao Y, et al. SPOCK1 as a potential cancer prognostic marker promotes the proliferation and metastasis of gallbladder cancer cells by activating the PI3K/AKT pathway. Molecular cancer. 2015;14(1):12.
162. Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. Current opinion in structural biology. 2001;11(1):39-46.
163. Tovar H, Garcia-Herrera R, Espinal-Enriquez J, Hernandez-Lemus E. Transcriptional master regulator analysis in breast cancer genetic networks. Computational biology and chemistry. 2015;59 Pt B:67-77.
164. Abildgaard MO, Borre M, Mortensen MM, Ulhoi BP, Torring N, Wild P, et al. Downregulation of zinc finger protein 132 in prostate cancer is associated with aberrant promoter hypermethylation and poor prognosis. International journal of cancer. 2012;130(4):885-95.
165. Wang Q, Curran ME, Splawski I, Burn TC, Millholland JM, VanRaay TJ, et al. Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. Nature genetics. 1996;12(1):17-23.
166. Nie L. Mutations of KCNQ4 Channels Associated with Nonsyndromic Progressive Sensorineural Hearing Loss. Current opinion in otolaryngology & head and neck surgery. 2008;16(5):441-4.
167. Soh H, Pant R, LoTurco JJ, Tzingounis AV. Conditional Deletions of Epilepsy-Associated KCNQ2 and KCNQ3 Channels from Cerebral Cortex Cause Differential Effects on Neuronal Excitability. The Journal of Neuroscience. 2014;34(15):5311-21.
168. Lehman A, Thouta S, Mancini GMS, Naidu S, van Slegtenhorst M, McWalter K, et al. Loss-of-Function and Gain-of-Function Mutations in KCNQ5 Cause Intellectual Disability or Epileptic Encephalopathy. Am J Hum Genet. 2017;101(1):65-74.
169. Burkhard P, Stetefeld J, Strelkov SV. Coiled coils: a highly versatile protein folding motif. Trends in cell biology. 2001;11(2):82-8.
170. Dai C, Tang Y, Jung SY, Qin J, Aaronson SA, Gu W. Differential effects on p53-mediated cell cycle arrest vs. apoptosis by p90. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(47):18937-42.
171. Zhao P, Guan HT, Dai ZJ, Ma YG, Liu XX, Wang XJ. Knockdown of SPOCK1 Inhibits the Proliferation and Invasion in Colorectal Cancer Cells by Suppressing the PI3K/Akt Pathway. Oncology research. 2016;24(6):437-45.
172. Zhang J, Zhi X, Shi S, Tao R, Chen P, Sun S, et al. SPOCK1 is up-regulated and promotes tumor growth via the PI3K/AKT signaling pathway in colorectal cancer. Biochemical and biophysical research communications. 2017;482(4):870-6.
173. Fan LC, Jeng YM, Lu YT, Lien HC. SPOCK1 Is a Novel Transforming Growth Factor-beta-Induced Myoepithelial Marker That Enhances Invasion and Correlates with Poor Prognosis in Breast Cancer. PloS one. 2016;11(9):e0162933.
174. Miao L, Wang Y, Xia H, Yao C, Cai H, Song Y. SPOCK1 is a novel transforming growth factor-beta target gene that regulates lung cancer cell epithelial-mesenchymal transition. Biochemical and biophysical research communications. 2013;440(4):792-7.
175. Li Y, Chen L, Chan TH, Liu M, Kong KL, Qiu JL, et al. SPOCK1 is regulated by CHD1L and blocks apoptosis and promotes HCC cell invasiveness and metastasis in mice. Gastroenterology. 2013;144(1):179-91.e4.

176. Martini M, De Santis MC, Braccini L, Gulluni F, Hirsch E. PI3K/AKT signaling pathway and cancer: an updated review. Annals of medicine. 2014;46(6):372-83.
177. Yang J, Yang Q, Yu J, Li X, Yu S, Zhang X. SPOCK1 promotes the proliferation, migration and invasion of glioma cells through PI3K/AKT and Wnt/beta-catenin signaling pathways. Oncology reports. 2016;35(6):3566-76.
178. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. Nature reviews Cancer. 2013;13(2):97-110.
179. Bates RC, Mercurio AM. The epithelial-mesenchymal transition (EMT) and colorectal cancer progression. Cancer biology & therapy. 2005;4(4):365-70.
180. Heller G, Altenberger C, Schmid B, Marhold M, Tomasich E, Ziegler B, et al. DNA methylation transcriptionally regulates the putative tumor cell growth suppressor ZNF677 in non-small cell lung cancers. Oncotarget. 2015;6(1):394-408.
181. Ashktorab H, Rahi H, Wansley D, Varma S, Shokrani B, Lee E, et al. Toward a comprehensive and systematic methylome signature in colorectal cancers. Epigenetics. 2013;8(8):807-15.
182. Pangeni RP, Channathodiyil P, Huen DS, Eagles LW, Johal BK, Pasha D, et al. The GALNT9, BNC1 and CCDC8 genes are frequently epigenetically dysregulated in breast tumours that metastasise to the brain. Clinical epigenetics. 2015;7:57.
183. Morris MR, Ricketts CJ, Gentle D, McRonald F, Carli N, Khalili H, et al. Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. Oncogene. 2011;30(12):1390-401.
184. Jiang GY, Zhang XP, Zhang Y, Xu HT, Wang L, Li QC, et al. Coiled-coil domain-containing protein 8 inhibits the invasiveness and migration of non-small cell lung cancer cells. Human pathology. 2016;56:64-73.
185. Yang CK, Yu TD, Han CY, Qin W, Liao XW, Yu L, et al. Genome-Wide Association Study of MKI67 Expression and its Clinical Implications in HBV-Related Hepatocellular Carcinoma in Southern China. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology. 2017;42(4):1342-57.
186. Garrigos C, Espinosa M, Salinas A, Osman I, Medina R, Taron M, et al. Single nucleotide polymorphisms as prognostic and predictive biomarkers in renal cell carcinoma. Oncotarget. 2017;8(63):106551-64.
187. Zhou G, Sun G, Zhou Y, Wang Q. Transcriptomic analysis of human non-small lung cancer cells A549 treated by one synthetic curcumin derivative MHMD. Cellular and molecular biology (Noisy-le-Grand, France). 2017;63(9):35-9.
188. Chen Q, Yao YT, Xu H, Chen YB, Gu M, Cai ZK, et al. SPOCK1 promotes tumor growth and metastasis in human prostate cancer. Drug design, development and therapy. 2016;10:2311-21.
189. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biology. 2016;5(1):e3.
190. Tournier B, Chapusot C, Courcet E, Martin L, Lepage C, Faivre J, et al. Why do results conflict regarding the prognostic value of the methylation status in colon cancers? The role of the preservation method. BMC cancer. 2012;12:12.
191. Irahara N, Nosho K, Baba Y, Shima K, Lindeman NI, Hazra A, et al. Precision of Pyrosequencing Assay to Measure LINE-1 Methylation in Colon Cancer, Normal Colonic Mucosa, and Peripheral Blood Cells. The Journal of molecular diagnostics : JMD. 2010;12(2):177-83.
192. Leong KJ, James J, Wen K, Taniere P, Morton DG, Bach SP, et al. Impact of tissue processing, archiving and enrichment techniques on DNA methylation yield in rectal carcinoma. Experimental and molecular pathology. 2013;95(3):343-9.

193. Jass JR. Molecular heterogeneity of colorectal cancer: Implications for cancer control. Surgical oncology. 2007;16 Suppl 1:S7-9.
194. Wang Q, Jia P, Cheng F, Zhao Z. Heterogeneous DNA methylation contributes to tumorigenesis through inducing the loss of coexpression connectivity in colorectal cancer. Genes, chromosomes & cancer. 2015;54(2):110-21.
195. Ooki A, Yamashita K, Kikuchi S, Sakuramoto S, Katada N, Kokubo K, et al. Potential utility of HOP homeobox gene promoter methylation as a marker of tumor aggressiveness in gastric cancer. Oncogene. 2010;29(22):3263-75.
196. Graham TA, McDonald SA, Wright NA. Field cancerization in the GI tract. Future oncology (London, England). 2011;7(8):981-93.
197. Luo Y, Yu M, Grady WM. Field cancerization in the colon: a role for aberrant DNA methylation? Gastroenterology report. 2014;2(1):16-20.
198. Shen L, Kondo Y, Hamilton SR, Rashid A, Issa JP. P14 methylation in human colon cancer is associated with microsatellite instability and wild-type p53. Gastroenterology. 2003;124(3):626-33.
199. Grady WM, Parkin RK, Mitchell PS, Lee JH, Kim YH, Tsuchiya KD, et al. Epigenetic silencing of the intronic microRNA hsa-miR-342 and its host gene EVL in colorectal cancer. Oncogene. 2008;27(27):3880-8.
200. Shen L, Kondo Y, Rosner GL, Xiao L, Hernandez NS, Vilaythong J, et al. MGMT promoter methylation and field defect in sporadic colorectal cancer. Journal of the National Cancer Institute. 2005;97(18):1330-8.
201. Wibe A, Rendedal PR, Svensson E, Norstein J, Eide TJ, Myrvold HE, et al. Prognostic significance of the circumferential resection margin following total mesorectal excision for rectal cancer. The British journal of surgery. 2002;89(3):327-34.
202. Sauer R, Becker H, Hohenberger W, Rodel C, Wittekind C, Fietkau R, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. The New England journal of medicine. 2004;351(17):1731-40.
203. Sauer R, Liersch T, Merkel S, Fietkau R, Hohenberger W, Hess C, et al. Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30(16):1926-33.
204. Bosset JF, Calais G, Mineur L, Maingon P, Stojanovic-Rundic S, Bensadoun RJ, et al. Fluorouracil-based adjuvant chemotherapy after preoperative chemoradiotherapy in rectal cancer: long-term results of the EORTC 22921 randomised study. The Lancet Oncology. 2014;15(2):184-90.
205. Ryan JE, Warrier SK, Lynch AC, Ramsay RG, Phillips WA, Heriot AG. Predicting pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review. Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland. 2016;18(3):234-46.