# Methylation changes in colitis associated cancer and its correlation to SMAD7

Pramodh Chitral Chandrasinghe

Division of Cancer

Department of Surgery and Cancer
Imperial College London

Thesis submitted for the degree of
Doctor of Philosophy

**Supervisors:**

Prof. Justin Stebbing
Mr. Janindra Warusavitarne
Dr. Biancastella Cereser

**DISCLAIMER**

I, Pramodh Chandrasinghe, declare that all the research presented in the thesis is my own work unless otherwise stated. Contributions from others are clearly stated and fully acknowledged.

**Abstract**

Ulcerative colitis (UC) is an inflammatory condition affecting the mucosa of the large bowel and the occurrence is linked to the increase of developing colitis associated cancer (CAC). The identification of biomarkers to predict those who are at risk of progressing to CAC is needed to assist in early decision making for restorative surgery and improved outcome following surgery. However, previous attempts at the identification of consistent genetic biomarkers have not been successful, due to non-reproducibility of the results. This could attribute partly to variability in tissues used and partly to the complex pathophysiology. The regenerative stress in UC mucosa resulting in multiple structural mutations in genes may also be a cause for the variability.

Therefore, the first aim of this study was the identification of differentially methylated genes to be used as potential biomarkers of malignant transformation at a precancerous stage. Whole genome bisulphite sequencing was carried out on laser captured epithelial cells on a pilot cohort of 8 samples of normal, inflamed, dysplastic and malignant colonic mucosa. Each sample was matched with adjacent non-neoplastic mucosa or buffy coat from the same patient as controls. Sixty-three hypomethylated and six hypermethylated gene promoters were identified as differentially methylated in the samples compared to the normal epithelium. These methylation changes were unique to the diseased tissue and were consistently found in each stage of the disease, suggesting they could be used as biomarkers for CAC. Out of these genes, 7 hypomethylated and 4 hypermethylated genes were identified as strongly related to CRC pathways. Similarly, the analysis of gene body methylation allowed the identification of further 10 hypomethylated and 2 hypermethylated genes related to CRC pathways. Moreover the identified genes

could be annotated to cell adhesion related molecular function and disease processes, including UC and epithelial cancers.

Amongst proteins and pathways of clinical importance, TGFβ, a molecule with anti-inflammatory properties, is inhibited in UC-affected mucosa. Recently, SMAD7, the principal intracellular inhibitor of the TGFβ pathway, has been proposed as a therapeutic target in UC. However, SMAD7 is shown to influence CRC development and progression in a less understood manner. Furthermore, studies in breast cancer indicate that SMAD7 influences methylation patterns of cancer-associated genes. Therefore investigating the potential effect of SMAD7 on CAC and differential methylation of related genes would provide insight into its role in CAC. The second and third aims of this study were therefore was to observe the behaviour pattern of SMAD7 at different stages of CAC and analyse the methylation pattern of identified SMAD7-associated genes.

Immunohistochemistry and *in situ* hybridisation was performed on 53 and 33 samples from non-inflamed, inflamed, dysplastic and cancer tissues, respectively. SMAD7 was biphasically expressed in different stages of CAC: high expression during the inflammation and cancer stages was associated with a lower expression in the non-inflamed UC and dysplastic tissues. On the other hand, the downstream molecule pSMAD3 did not show a reduction in expression suggesting an escape of TGFβ from the inhibitory effect of SMAD7 in UC. The evaluation of the methylation patterns in cancer-associated genes supposedly influenced by SMAD7 (*CDH1, CLDN4, DNMT1, CGN*) did not demonstrate a consistent pattern of differential methylation. This data has strong clinical implications, as the biphasic expression of SMAD7 during precancerous stages suggests further evaluation is needed prior to using an antisense to inhibit SMAD7 in the clinic.

In conclusion, this pilot study identified 27 potential genes with differential methylation, of which the expression of some has already been validated with RT-qPCR. The consistency of the methylation pattern of these genes across all stages of CAC and their annotation to cancer and cancer related biological processes make them strong potential candidates for an early biomarker in CAC. An extension of the analysis in a wider samples cohort and complete validation of each target will be necessary to validate their efficacy and lead to developing a clinically useful gene panel to predict CAC.

## Acknowledgements

I wish to thank my supervisors for standing by me, encouraging and guiding me to complete this work. Mr Janindra Warusavitarne as my clinical trainer, encouraged me to take up a research degree and was the pillar of strength behind my achievement. Special thanks to Prof. Justin Stebbing for making my dream a reality by providing me the opportunity to join Imperial College under his supervision and guiding me to new heights. My sincere gratitude to Dr Biancastella Cereser who supervised me, taught me and guided me all throughout the past four years. She was a continuous pillar of strength standing by me at all times. Special thanks to my wife Laknadhi, my two daughters, Hesanga and Dinanga, for all the support and tolerating my long absence from home. Without their support it would have been impossible to achieve this great feat. I wish to thank everyone who worked at the 'Stebbing Lab' for their kind manner and support given to me throughout the years. A special thanks to Dr Lisa Del Bel Belluz (Research technician), Dr Philip Carter (Bioinformatician), Dr Angela Yiu (former PhD student), Dr Nina Moderau (Research Associate) and Dr Sladana Zagorac (Research Associate) for the assistance and guidance they provided me during difficult times. I also thank my colleagues Neha Tabassum, Jinrui Wen and Sandesh Meyler for support and companionship they provided.

Dr Morgan Moorgen, Consultant Pathologist from St. Mark's Hospital, London was instrumental in the success of my project by providing the access to tissue samples and histopathological assessment.

The Commonwealth Scholarship Commission was the funding agency for my PhD without whom my dream of completing a doctoral degree at Imperial would have been a dream. My sincere thanks to all the officials at the Commonwealth who enabled me to continue my studies smoothly throughout the years. Lastly a huge

appreciation to all the patients who agreed to participate and provide the tissue samples, without whom my work would not be possible.

*Dedicated to*

*my family,*

*parents*

*and*

*all the patients who taught me throughout the years…*

## Table of Contents

## List of figures

## List of tables

## List of abbreviations

BMP – Bone morphogenic protein

CAC – Colitis associated cancer

CCO - Cytochrome c oxidase

CD – Crohn's disease

CGI - CpG island

CRC – colorectal cancer

CRISPR - clustered regularly interspaced short palindromic repeats

DAB - Diaminobenzidine tetra hydrochloride

DMR – Differentially methylated regions

DNMT - DNA methyl transferase

EMT – Epithelial mesenchymal transmission

FAP – Familial adenomatous polyposis coli

FFPE - Formalin fixed paraffin embedded

GAPDH - Glyceraldehyde 3- phosphate dehydrogenase

GO – Gene ontology

H&E - Haematoxylin and Eosin

HNPCC - Hereditary non polyposis colorectal cancer

HRA - Health Research Authority

IBD – Inflammatory bowel disease

ISH – *In situ* hybridization

LCM - Laser capture microdissection

MAP – *MUTYH* associated adenomatous polyposis

OCT – optimal cutting temperature

PBS - Phosphate buffered saline

PCR - Polymerase chain reaction

PLC - programmable logic controllers

PPIB - Peptidyl propyl isomerase B

PPIN – Protein –protein interaction network

RT-qPCR - Real-time polymerase chain reaction

SMAD – Small mothers against decapentaplegic

TGF-β Transforming growth factor β

UC – Ulcerative colitis

WGBS – Whole genome bisulphite sequencing

# 1.Introduction

## 1.1 Colorectal cancer

Colorectal cancer (CRC) is the 3rd leading cause of death due to malignancies globally (1). Traditionally thought to be a disease of the industrialised communities, currently there is a surge of CRC incidence in the low- and middle-income earning countries (1, 2). The majority of CRC are sporadic cancers while 15 to 20% occur due to identifiable predisposing conditions. Amongst these are the hereditary polyposis conditions such as Familial adenomatous polyposis (FAP), Attenuated FAP, MUTYH associated polyposis syndrome (MAP) and Hereditary non polyposis colorectal cancer (HNPCC) (3). There are other rare hereditary syndromic conditions, such as Peutz Jeggar's, Gardner's, Cowden and Turcot syndrome, which cause CRC. Inflammatory bowel disease (IBD), comprising of Ulcerative Colitis (UC) and Crohn's disease (CD), is an independent risk factor for CRC. However, while UC has a well-established link to increased risk of CRC, the risk due to CD is controversial (4). Survival from CRC has improved over the recent past due to the improvement in surveillance methods, surgical techniques and chemotherapy protocols. Risk reduction has also been effective due to health education and population screening for premalignant stages. However, IBD being a non-modifiable risk factor and its incidence increasing, UC related CRC incidence has a potential to increase in the future.

## 1.2 Ulcerative colitis

Ulcerative colitis is an inflammatory condition of autoimmune nature affecting the mucosa of the large bowel. The condition is part of the inflammatory bowel disease group where Crohn's disease and indeterminate colitis are the other counter parts. It is estimated that each year, in Europe, new cases are diagnosed at a rate of

between 0.9 to 24.3 per 100,000 person (5). There is a difference in the incidence between the west and the east of the continent, with the highest annual incidence of 24.3 recorded in Iceland and the lowest rates in Eastern Europe. Worldwide, it has also been noted that the disease occurrence increases proportionately to the human development index, where a higher incidence and prevalence of the disease is observed in the industrialised populations (6-9).

### 1.2.1 Pathophysiology of UC

The exact pathophysiology of UC is not known, although it is thought to be of multifactorial nature (10-12). Aetiology of the disease has been linked to exposure of the immune system to antigens during the early stage of life (13-15). Emerging data suggests a connection between UC and the gut microbiome (16). The underlying pathophysiology appears to be a deregulated mucosal immune response driven by genetic susceptibility and gut commensals (17). Inflammation in UC is confined to the mucosa, in contrast to the transmural inflammation seen in Crohn's disease. In the normal colon, the epithelial cells with their tight junctions and mucus layer separate the luminal commensals from the immune cells in the lamina propria. Individuals with UC have defective mucus secretion and week tight junctions, which would allow the antigens to come in to contact with the immune cells (18, 19). However, whether the deranged defences are a cause or the effect of inflammation is yet unclear (17). The antigen presenting cells (dendritic cells) increase in the lamina propria of patients with UC resulting in activation of NF-kB and migration of T cells. There is an atypical T helper type 2 response that is mediated by non-classic natural killer T-cells producing interleukins 5 and 13. The interleukins have a cytotoxic effect causing an

inflammation in the mucosal layer (19, 20). An exaggerated leucocyte response is brought about by a combination of interleukins and cytokines such as Tumour necrosis factor (TNF)-α resulting in a chronic inflammatory response.

There are well-recognised genetic markers that predict the susceptibility for UC. The major histocompatibility complex class-2 region near HLA-DRA has shown strong association with IBD while epithelial cell adhesion molecules such as hepatocyte nuclear factor-4α, E-cadherin1 (CDH1), and laminin-β1 have been specific for UC (21). Methylation of *CDH1* was the first association between UC and CRC that was recognised (22). Using methylation specific PCR on case matched samples, this study demonstrated methylated promoter of *CDH1* in 57% of the UC associated CRC samples and 36% of the sporadic CRC. After confirming the authors concluded that there is significant overlap between the two types of cancers in genetic pathways. This may indicate an important role of early methylation changes in CRC that is independent of inflammation. As the scientific thinking behind bowel cancer on UC is changing from that of oxidative stress to regenerative stress (23) these early epigenetic markers might hold the answer for a risk predictor of CAC.

### 1.2.2 Clinical management of UC

Currently patients with active disease are monitored periodically with colonoscopy and biopsy of suspected areas of the mucosa (24). Those who have dysplastic changes are offered surgery to remove the colon, which will be replaced either with an ileal pouch (restorative surgery) or a permanent ileostomy (25-28). As I will discuss further, high-grade inflammation is an absolute risk of development of cancer warranting the removal of the colon (29). However, the colonoscopic evaluation is operator-dependent and in some series around 30% of those who undergo surgery

for dysplasia has been found to harbour occult cancer in the lesions on histopathological examination (30, 31). *These factors necessitate the search for biomarkers predictive of progression to carcinoma*. Early referral to surgery can prevent the development of cancer and improve the outcome from restorative surgery (25, 27).

### 1.3 UC associated CRC

While UC initially presents as blood and mucus diarrhoea, it can progress to life threatening conditions such as toxic megacolon and perforation of the colon in the acute stage (32), and cancer. Patients with long standing UC have in fact an increased risk of developing colitis-associated colorectal cancer (CAC) (33-37), a risk that is proportionate to the extent and the duration of the disease (30, 31, 38). Traditionally, the risk of developing colitis-associated colorectal cancer CAC is believed to be 2% by 10 years, 8% by 20 years, and 18% by 30 years duration of UC (33, 39). Although modern surveillance has reduced the risk of CAC, population based large scale studies indicate UC patients are characterised by an increased risk of developing caners in the large bowel compared to the normal population (40-42). A recent study indicates that the risk of colorectal neoplasia in UC doubles for every 10 years of mild, 5 years of moderate or 3.3 years of severe active disease, and reject the conventional belief of neoplastic risk being related only to the duration of disease (30). The same study suggests that the stress of the cumulative inflammatory burden on the colonic mucosa to be causing the malignant transformation as previously observed (31).

Many attempts have been made at recognising unique genetic mutations in CAC to be used to risk stratify (22, 23, 43-50). However, none of the studies have been able to identify a single or a group of candidate gene mutations that could be used as risk predictors. Fort this reason, the focus has recently shifted to looking into methylation markers (44, 47, 51-62). A risk predictor, to be effective, needs to be present in early stages (pre-cancerous) and directly related to the process of CRC development. A consistent methylation marker that precedes the malignant transformation of colonic epithelial cells can be an effective risk predictor. If it is demonstrated in colonoscopic biopsy samples early decision-making regarding surgical removal of the diseased colon can be facilitated.

## 1.4 The mutational landscape of CAC

There are considerable differences between CAC and sporadic CRC (38, 63, 64). A sporadic CRC implies a large bowel cancer occurring in an individual without an identifiable high-risk condition such as inflammation polyposis syndromes or a strong family history. CAC refers to large bowel cancers that arise in the background of long-standing inflammatory conditions such as UC and CD. While sporadic CRC starts from a well-studied sequence of changes from normal epithelium to adenoma and then cancer, (65) CAC progresses from inflammation through dysplasia to cancer (66). The lesions in CAC are flat field changes making the identification difficult, as opposed to elevated polyps which are present in sporadic cancer (67). The progression from inflammation to dysplasia in UC-patients is also difficult to be predicted on clinical grounds. The rate of progression to neoplasia from inflammation is variable while in some case series a rate of over 50% progression to neoplasia has

been reported (63, 68, 69). The progression from dysplasia to CAC is also time dependent and unpredictable. Active surveillance of early neoplastic disease has not led to early detection or prevention of progression to cancer. Currently serial surveillance endoscopy is the preferred method to search for progression to CAC. It is both invasive and costly. Moreover, the lack of precision of endoscopic surveillance is also reported with unsuspected neoplasia or malignancy in resected specimens (70). These issues all highlight the necessity of an early molecular marker that can be detected at the inflammatory stage, enabling early decision making for colectomy.

The progression to CAC from UC occurs in a stepwise process on colonic epithelium that has field cancerisation effect in contrast to 'adenoma-carcinoma sequence' in sporadic where cancer arises on a benign polypoidal growth (**Figure 1**). The initial active chronic inflammatory stage is followed by dysplastic changes in the mucosa, defined as morphological changes in the epithelial cells without invasion through the basement membrane. Dysplasia of low grade can then progress to a high-grade dysplasia and eventually lead to cancer. The progression is unpredictable, although the risk of cancer exponentially increases after 8 to 10 years from the onset of the UC (71).

Evidence so far indicates that the progression to cancer is due to the on going chronic inflammation of the colon, which causes DNA damage from oxidative stress (72, 73). However, new research link the presence of mutations caused by regenerative stress to the malignant transformation (23). The inflammation resulting in an increased cell turnover leading to increased replicative burden and higher chances of cancer causing mutations is suggested (23).

The mutational landscape of UC differs from that of sporadic cancer (23, 64, 66, 74), (**Figure 2**). The mutation in Adenomatous Polyposis Coli (*APC*) gene, present as an early and the commonest mutation in sporadic cancer, is seen in around 15% of CAC as a late event (43). On the other hand, mutations in both Tumour Protein p53 *(TP53*) which occur late in CRC, and Kirsten-Rat Sarcoma (*KRAS*) proto-oncogene are detected much earlier in the course of malignant transformation in CAC, and can be considered initiating mutations of this neoplasia (46, 48). As mentioned above, multiple studies are yet to identify unique genetic markers specially to predict the malignant transformation in premalignant stages (43, 46-49, 75).

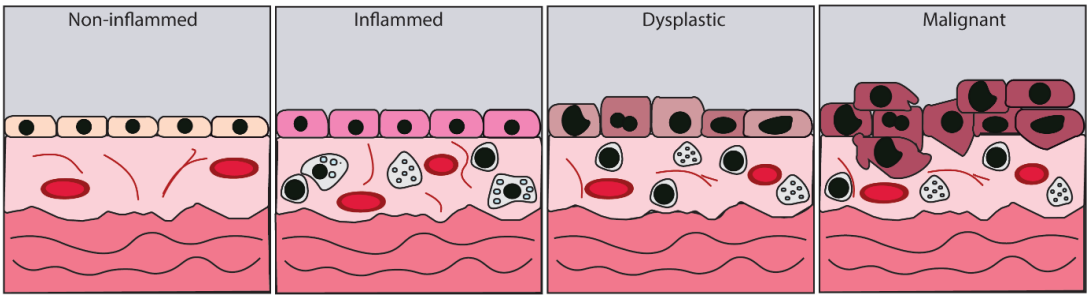



**Figure 1. Stages of colitis-associated cancer (CAC).**

Schematic representation of cross sections of the colon showing progression of the colonic mucosa through different stages of colitis associated neoplasia. Infiltration of the submucosal tissue with inflammatory cells occurs in the inflammatory stage followed by non-invasive cellular and nuclear changes in the mucosa during dysplastic phase. The transformed cells in the malignant stage eventually invade into the submucosal layers in cancer.

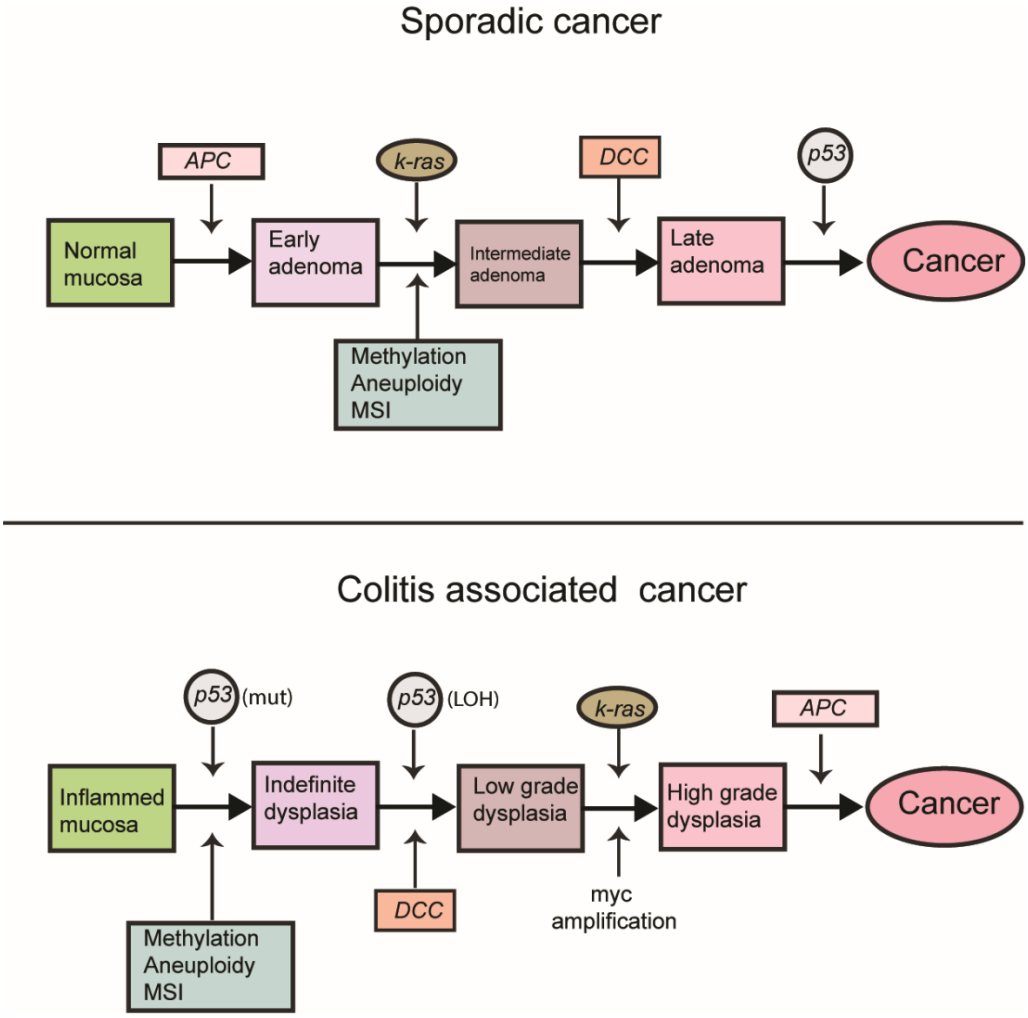



**Figure 2. Molecular progression to cancer in sporadic CRC and CAC.**

The sequence of genetic mutations in the two pathways differs. The most contrasting are the mutations of *APC* and *P53* genes, which occur at different, time points in the two pathways. Reproduced from Chandrasinghe et al 2018 (76).

## 1.5 CAC in relation to consensus molecular subtypes (CMS) of CRC

Colorectal cancer is subdivided into four molecular subtypes based on the genetic landscape of tumours (77-79). CMS 1 is characterised by a high CpG island methylation phenotype (CIMP), high expression of mutated BRAF V600E and impaired DNA mismatch repair (MMR) function. CMS 2, which is the most common subtype, includes the classical adenoma-carcinoma sequence of CRC development with the early loss of *APC* gene, activation of k-RAS mutation and late loss of P53 function. CMS 3 shows features consistent with chromosomal instability (CIN). It also harbours more microsatellite instability (MSI) and has the highest rates of k-RAS mutations. CMS 4 subtype is related to pro-inflammatory tumour microenvironment, with higher somatic copy number alteration (SCNA) and high stromal involvement. CMS 4 has the worst prognosis out of all subtypes, with the presentation at an advanced stage and a greater chance of dissemination. Although CAC has a molecular profile, which overlaps on several subtypes (23, 48, 50, 80) and it has not been included in this classification applicable to sporadic CRC, it has been linked to CMS 4 based on several molecular characteristics. The presence of pro-inflammatory cytokines IL-23 and IL-17 in the CMS 4 subtype is infact common to both sporadic CRC and CAC. Although P53 mutations in CAC are abundant as in CMS 2, the mutations are seen much earlier in the process as opposed to a later development on CMS 2. Epithelial-mesenchymal transition (EMT) is another feature shared between CAC with CMS 4. The stroma of CMS 4 cancers are infiltrated with epithelial cells, carcinoma associated fibroblasts (CAF) and innate immune cells (78), similar to CAC, where EMT has been well demonstrated (81). The highly aggressive

behaviour seen in CAC has been linked to the immune evading tumour microenvironment increasing chances of local invasion and dissemination (82, 83). Finally, higher rates of SCNA have also been demonstrated as a marker of CAC, which further strengthens the link between it and CMS 4 (84-86).

## 1.6 Prognosis in CAC compared to sporadic CRC

CAC has a higher tendency to manifest as multifocal cancer and harbour unfavourable histological features such as mucinous or signet cell variant which are associated with poor prognosis (87-89). Most of the earlier survival studies have reported a poorer outcome in CAC compared to sporadic CRC (90-93). Lavery et al observed that the survival when compared stage wise was comparable with that of sporadic cancer . They attributed the poor overall survival due to a higher percentage of CAC presenting at an advanced stage (91). Arnio et al comparing CRC in polyposis syndromes, sporadic cancers and CAC observed the worst outcome in CAC with a 28% 5 year survival rate (92). However more recent studies comparing CAC with sporadic CRC have reported comparable outcomes (94, 95). These studies are mostly single centre studies with a small sample size, which may imply a selection bias. A recent cancer register based study analysing data from over 67000 patients reported have reported a worse outcome in patients with CAC compared to sporadic CRC (96). In a subgroup analysis, this poorer prognosis is shown to be driven by the cohort of younger patients (less than 65 years). This is in contrast to an earlier population registry based Dutch study that reported worse prognosis in older patients with CAC (97). There are several hypotheses to explain the poor survival in CAC as observed by some authors. The multifocal nature of CAC arising in a background of field cancerisation along with an enhanced aggressive behaviour

of CAC and advanced stage at presentation are postulated to be resulting in poor outcome. Improvement in screening colonoscopy facilities has shown to increase detection of early stage cancers and precancerous stages resulting in a better survival (98-100). However the interesting phenomenon of interval cancers remain to be unexplained. Interval cancers are those cancers that appear between two surveillance colonoscopies and these cancers vary from early to advanced stage at presentation (98). These are probably the results of dysplastic lesions missed during surveillance despite of the tedious methods applied. A predictive molecular marker of CAC identified at early stages of inflammation enabling surgical extirpation of the colon may assist to prevent these interval cancers. Existing evidence suggest that there is an improvement in prognosis attributed to improved surveillance enabling early detection. However there are yet unexplained phenomena such as interval cancers and unpredictable aggressiveness of CAC resulting in poor outcome that require better understanding at a molecular level.

.

## 1.7 DNA methylation

Epigenetic changes are heritable phenotypic changes that take place in the genomic material without a change in the nucleotide sequence (101, 102).

DNA methylation involves the transfer of a methyl ($CH_3$) group to the $5^{th}$ carbon of the cytosine, forming 5-methyl cytosine (5mC). DNA methyl transferase (DNMT) enzymes catalyse this reaction (103). DNMT1 catalyses methylation during DNA replication while DNMT3A and DNMT3B catalyse *de novo* methylation (103).

The majority of methylation occurs at cytosine residues that are preceding a guanine nucleotide (CpG sites) (104-106). Across the genome, 70-80% of the CpG sites are

methylated except for CpG islands (CGIs) that are regions of around 1kb size with repeating CG sequences, protected from stochastic methylation in regulated genes (107-109). Most of these CGIs are promoters are at the 5' end of genes, and methylation in these regions influences gene transcription (110). Methylation is a dynamic process and can be reversed. Demethylation occurs as a passive process mediated by TET enzymes, which prevent DNMT1 from transferring the methylation to replicated DNA (111). A recently identified 5 hydroxymethyl cytosine (5hmC) is an intermediary state during demethylation of 5mC (112).

As mentioned above, methylation/demethylation in the CGIs contributes to the regulation of gene expression (110, 111) in several systems. For example, methylation of one X chromosome in females contributes to its inactivation, and methylation is shown to influence genetic imprinting, foetal development and tumorigenesis (104, 113).

Several possible mechanisms by which methylation regulate gene expression have been postulated:

1. Interference with binding of transcription factors

   Binding of transcription factors to the promoter region which is essential for gene transcription is prevented by methylation of CGIs (114).

2. Recruitment of repressor molecules

   Recruitment of specific molecules that bind the methylated CpG sequences in promoter region inhibits gene transcription (115).

### 1.7.1 Methylation changes in inflammatory bowel disease

Several studies have shown the altered regulation of transcription in IBD (56, 116, 117); however, there is to date no consensus on a list of hypo- or hypermethylated genes which characterise any stage of IBD, and in particular UC. A susceptibility locus on *DNMT3A* has been identified (118) and several authors have reported on differential methylation in inflamed colonic mucosa (56, 59, 61, 119).

The lack of progress in this particular field may be due to the fact that most studies have used mucosal biopsies instead of purified epithelial cells, which may affect the results due to the heterogeneity of epithelial, mesenchymal and immune cells included in these samples. Few studies have attempted to solve the confounding factor of cellular heterogeneity by enriching or chemically separating epithelial cells (54-56). This strategy allowed the identification of 7 candidate differentially methylated genes (*THRAP2, FANCC, GBGT1, DOK2, TNFSF4, TNFSF12,* and *FUT7*) in a small cohort of 8 IBD (both UC and CD) patients (56). Interestingly, the authors also identified different sets of methylated genes in the epithelial and stromal components of the same samples, highlighting the importance of studying the components separately. A summary of the studies investigating methylation markers in IBD with the identified genes is listed below **(Table 1)**.

| Methylated gene | Samples studied | Reference |
|---|---|---|
| **VMP1, WDR8** | 3 | *Ventham, Kennedy (120)* |
| **KCNK4, ICAM3, C6orf25, CARD9, IL8RB, IL8RA, C6orf27, TNNI2, CDH1, CARD9, ADA, SMPD3** | 8 | *Cooke, Zhang (56)* |
| **PITX2, ROR1, GXYLT2, RARB, FOXA2** | 5 | *Barnicle, Seoighe (54)* |
| **TRIM39-RPP2, TRAF6** | 9 | *McDermott, Ryan (121)* |
| **CFI, FLNA, HKDC1, IGHG1, MT1H, PTN, SLC7A7b, SPINK4, THY1, TK1** | 30 | *Hasler, Feng (59)* |

**Table 1. Previously identified methylated genes in UC.**

List of differentially methylated genes identified in the inflammatory phase of UC based on literature. The table lists the genes described in literature by studying tissue samples from patients with UC. All studies have included tissue samples with both epithelial and stromal components. Lack of a consistent pattern is noted amongst the studies.

### 1.7.2 Methylation in colitis associated cancer (CAC)

Research aimed at the identification of methylation markers in CAC has not been extensively undertaken up to now. Epigenetic, rather than genetic, changes, have been suggested to result in better biomarkers to predict neoplasia in UC due to several practical reasons, including the fact that methylation changes are stable enough changes that can be studied in frozen and also in paraffin embedded tissue samples with modern sequencing methods (122), which was not possible until recently. Also, the methylation changes occur in longer segments of DNA than single nucleotide polymorphisms (SNP) and can be seen in a wider region of (field change) the affected tissue (123). Alike the research on the epigenetics of IBD, several attempts have been made at identifying methylation markers in CAC. Beggs et al studied samples from 4 patients with dysplasia and identified *TUBB6* as a gene differentially methylated in dysplasia associated with colitis (55). Emmett et al in a systematic review observed the differences in promoter methylation in CAC and sporadic CRC (57). They concluded and *ESR* as being highly methylated in CAC. Garrity-Park et al looked in to the non-neoplastic regions in patients with CAC and recognised few gens; *MINT1, RUNX3* and *COX-2* to be hypermethylated in the non-neoplastic regions of those with CAC (58).

Even though these studies have an impact on our knowledge of epigenetic changes in CAC, it is still unclear how methylation is differently regulated across all stages of CAC development. Identifying methylated genes from tissues in different stages of CAC (inflamed, dysplastic, malignant) may assist in guiding therapy in CAC but the predictive value will be in a set of genes that is related to malignant transformation and can be recognised at a precancerous stage.

Other than the identification of a specific biomarker for each stage of CAC, the study of epigenetic changes could also be clinically translatable with the deployment of drugs, which can target proteins or genes regulated by, or regulating methylation. An example of this translational potential can be seen in the use of an siRNA targeting SMAD7, (124-128) a protein expressed in the inflamed colonic mucosa, which is also involved in methylation processes in different cancers, as I will mention below.

## 1.8 SMAD7

Mothers against decapentaplegic (MAD) gene was first identified during a search for mutations in the decapentaplegic (*DPP*) gene which is responsible for the formation of *Drosophila* wings (129). It was later identified that three homologues of *MAD* in *C. elegans* are responsible for its small body size hence the name small mothers against decapentaplegic (SMAD) (130). SMAD proteins consist of two globular proteins (MH1 and MH2) joined by a non-structural linker (**Figure 3**). The hairpin structure with DNA binding capability is called the MH1 domain and its preference for specific base sequences differs amongst different SMADs (131). The hydrophobic MH2 domain binds to trans-membrane receptors TGFβR and BMPR.

Members of the TGFβ superfamily, including TGFβ1, bone morphogenic protein (BMP), Nodal and activin, exert their effect through phosphorylation of SMAD proteins (**Figure 4**). Transmembrane receptors with threonine kinase activity, TRβI and TRβII, phosphorylate the SMADs upon activation by the members of the TGFβ superfamily. Respectively, TGFβ1 pathway involves SMAD2 and SMAD3, while BMP pathway recruits SMAD1, SMAD5 and SMAD8 as a signal transducing complex (112). The final common pathway involves the binding of SMAD4 with the activated

complexes and the migration to the nucleus. This complex of SMADs can bind directly to the chromatin and regulate transcription. SMAD7, and SMAD6 to a lesser degree, antagonise both signalling pathways by inhibiting phosphorylation of the SMADs which prevents their nuclear migration (112). Ultimately, the SMAD complexes affect the transcription of several genes, including those related to carcinogenesis. In fact, both pro- and anti-carcinogenic pathways, such as *WNT, P53*, methylation changes in stem cell differentiation and expression of master transcription factors (*OCT4, SOX2, NANOG*) are influenced by the SMAD complexes (76, 132-134).

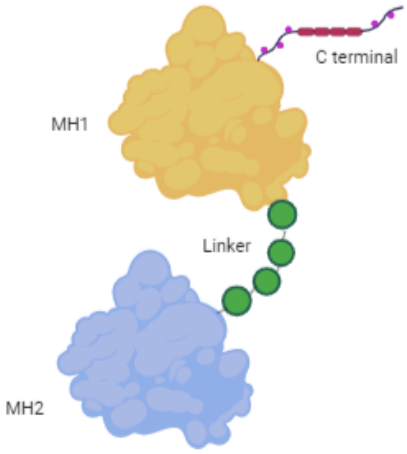



**Figure 3. Molecular structure of SMAD7.**

SMAD7 molecules comprise of two domains – MH1 and MH2 linked by a non-structural linker. The MH2 domain is hydrophobic and binds to the TGFβR receptor inhibiting its positive feedback loop.

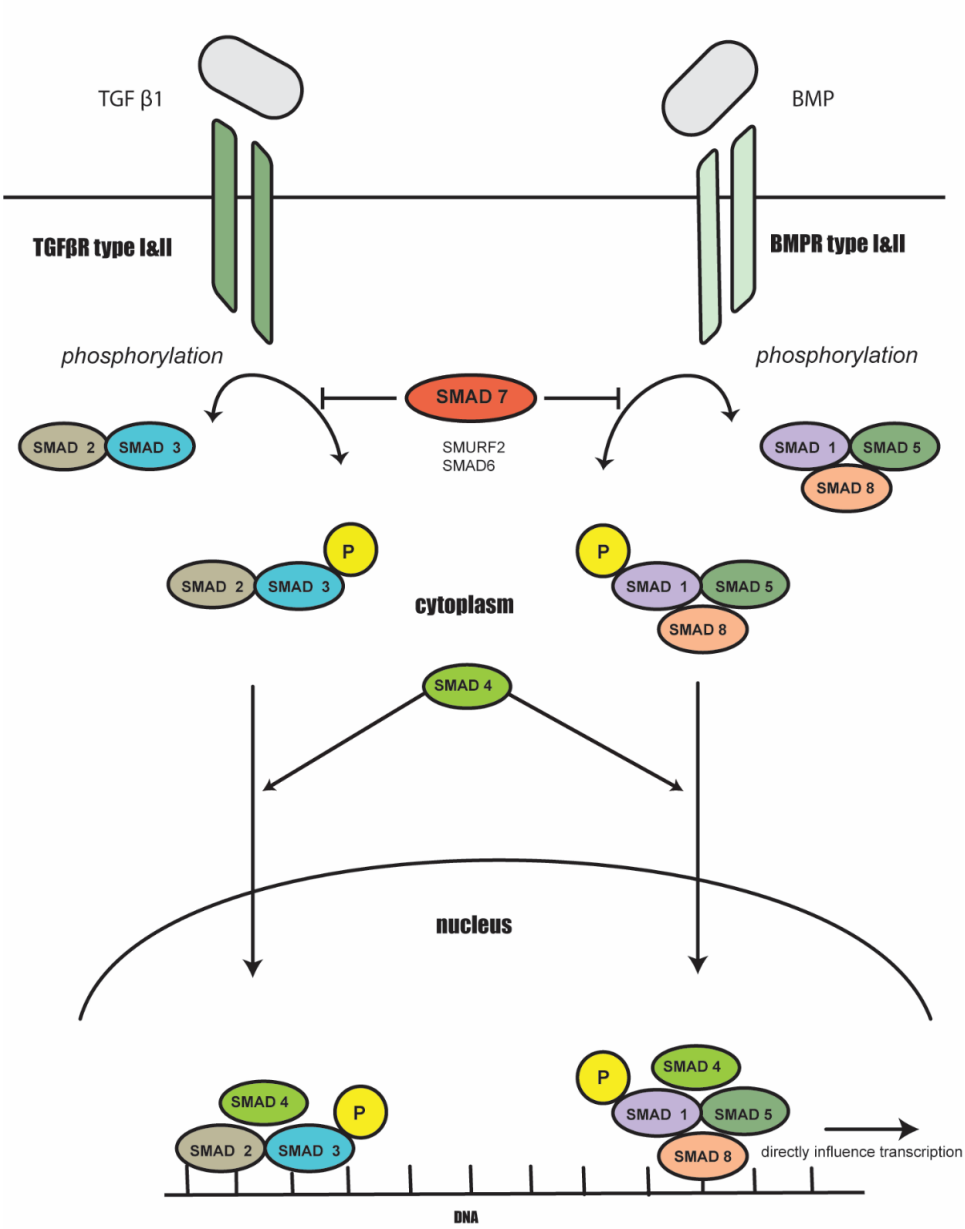



**Figure 4. The transforming growth factor β pathway**

SMAD7 inhibits the phosphorylation of SMAD2/3 and SMAD1/5/8 complexes blocking the intracellular signal transduction. The phosphorylation is required for the TGFβ signal to be transmitted to the nucleus in combination with SMAD4 and alter transcription of anti-inflammatory factors.

### 1.8.1 SMAD7 influencing gene methylation

An interesting study has demonstrated that overexpression of SMAD7 induces the hypomethylation of key cell membrane genes involved in epithelial mesenchymal transition (EMT) in breast cancer cell lines (133), such as *CDH1*, *CLDN4* and *CGN*. Therefore, SMAD7 in breast cancer may be protective towards preventing invasion and metastasis, as it prevents cell migration and tighter cell adhesion. To support these findings, inactivation of SMAD2 also results in the inhibition of DNMT1 causing hypomethylation of *CLDN1* in breast cancer cell lines (135). Furthermore, it has been demonstrated that the TGFβ pathway directly influences demethylation of tumour suppressor gene *CDKN2B* (which encodes $P15^{ink4b}$) in proliferating epithelial cells (136). Products of the oncogene *ZNF217* have shown to prevent the demethylation of *CDKN2B* by inhibiting the entry of SMAD2/3 complex into the nucleus. Thereby it inhibits the tumour suppressor activity of *CDKN2B*. Similarly, SMAD7, which inhibits the phosphorylation of SMAD2/3 by preventing the entry of the complex to the nucleus, has the same effect on CDKN2B. The cell cycle regulator protein FBOX32 has been found to have a lower expression due to gene methylation influenced by SMAD2/3 in oesophageal cancer (137). SMAD7 having a direct inhibit effect on SMAD2/3 complex can also influence this process. Similar epigenetic silencing of malignant cells by TGFβ has been demonstrated in ovarian cancer as well (138).

Ample evidence exists to suggest a direct influence of the TGFβ pathway and its components in the epigenetic modification of cancer-causing genes, which ultimately makes SMAD7, an inhibitory molecule of the pathway, a key regulator of this mechanism (139). The changes of SMAD7 in the UC affected colonic mucosa therefore may be playing a key role in forming CAC which has not been studied in detail.

### 1.8.2 Smad7 in UC

The mechanism underlying the initiating inflammation in UC is yet to be understood. SMAD7 has been long identified as a molecule that is highly expressed in UC affected colonic mucosa (124, 126, 140-142). Degradation of SMAD7 is delayed through acetylation of the molecule and it is hypothesised to sustains the inflammation by inhibiting the anti-inflammatory effect of TGFβ1 on the colonic mucosa (142). To substantiate this, it has been demonstrated that the inhibition of SMAD7 also resolves the inflammation in the colonic mucosa by restoring the TGFβ1 activity (140). These effects have been demonstrated in mouse models which later led to an attempt to inhibit SMAD7 as therapy for IBD by using a siRNA as an oral medication (127). The attempts have not been successful for Crohn's disease, and the data on UC is not yet available (143). However, the effect of regulating SMAD7 on the development of CAC has not been studied in detail. There has been equivocal evidence in few *in vitro* and murine studies done on sporadic and colitis associated cancers with regards to the effect of SMAD7 in development of CRC (125, 144, 145). High expression of SMAD7 in T cells of the UC affected mucosa increase the inflammation while demonstrating lower rates of developing CAC in mice (146). TGFβ1 has clearly shown to exert a dual effect on early and late stages of the carcinogenic process. During early transformation to cancer, TGFβ1 has a tumour suppressor action, while during the later stages of carcinogenesis TGFβ1 exerts a pro-carcinogenic effect (138, 147-151). Due to this, one could postulate that the modulation of TGFβ1 may affect the final outcome. For example, we can hypothesise that if SMAD7, an inhibitor of TGFβ1, is inhibited in a mucosa with dysplasia, it may induce a pro-carcinogenic action of TGFβ1 that is seen in late carcinogenesis.

### 1.9 Biological models available for the study of UC and CAC

Animal models with both chemical and transgenic induced colitis, have been utilised by researchers to study the aetiology and effect of UC. The commonly used chemical induction agents are dextran sodium sulphate (DSS), Oxazolone (OZ) and trinitrobenzene sulfonic acid (TNBS) (13). All agents induce a non-specific colitis upon oral or rectal administration. DSS causes injury to the epithelial barrier of the colon exposing the submucosal components to the bacteria and antigens in the lumen causing an inflammatory response (152). The colitis lasts for 5 to 7 days and pulse doses need to be administered to maintain a long-term inflammatory process to study the effects of UC. Although its technical simplicity has allowed its wide use, the relevance to human form of UC is questionable. The cause for epithelial injury in UC, which triggers the inflammatory response by exposing the antigens to immune cells, is yet to be identified. Therefore, a model that utilises an externally introduced toxic substance will have a confounding effect on the process and outcome. Additionally, it has been demonstrated that DSS-induced inflammation varies with the constitution of gut bacteriome. Mice treated with antibiotics have been affected by a lethal form of colitis, which further questions its applicability to study the human form of UC (153). Moreover, strains C57BL/6 shows a more pronounced effect on DSS than the BALB/c strain.

TNBS is a high molecular weight heptane protein, which induces a cell-mediated acute Th1 mediated inflammation upon administration. The immune response for this chemical reaction is mediated by NOD2, a receptor for bacterial antigen, which simulates the pathophysiology of CD more than UC (154). OZ when administered with ethanol into the rectum of mice has demonstrated a Th2 mediated immune response with marked elevation of IL-4 and 5 levels. This response more closely

resembles UC than CD (155). Compounds containing acetic acid have also been used to induce inflammation although their inherent risk of causing bowel perforation has caused caution (156). Although chemical models provide an environment to study the effects of longstanding inflammation in the colon, several confounding factors negate their scientific value. Since these compounds need to be administered in multiple doses to maintain the inflammation, the effect observed will be different to that of continuing inflammation. The effect of compounds such as acetic acid could alter the genetic material in the mucosal cells affecting the genetic data acquired through such models. Most of these compounds being acidic or basic and containing benzene compounds may have a profound effect on genetic and epigenetic changes on the mucosal cells (157, 158). In particular, benzene has been shown to alter the methylation in key neoplasia related genes such as Cdkn2b and Runx3 (159).

Transgenic mouse models have the advantage of being able to study isolated specific pathways in order to recognise the exact mechanism of inflammation. IL10 knockdown mouse models (IL-10-KO), a commonly used transgenic model, shows an inflammation upon exposure to the environment due to the absence of the anti-inflammatory effect of IL10 cytokine. The models also show an expansion of Th1 and Th17 cells and defective function of regulatory T cells, resulting in a chronic inflammatory reaction (160). The colonisation of the gut microbiome is required for this process, and mice in clean environment do not elicit an inflammatory reaction. This transgenic model renders researchers a more potentially similar environment to study the effects of UC due to its 'multihit' mode of initiation (genetic and environmental factors) and a predictable alteration in the immune pathways. MDR1a knockdown mice elicit colitis due to the disruption of the colonic epithelial barrier (161). Multidrug resistance pump in the colonic epithelium is deficient in these mice,

who develop colitis spontaneously. The advantage of this model is that it closely represents the current widely accepted hypothesis of epithelial barrier dysfunction as the aetiology of IBD (162). ITF-dnRII, SMAD-3 and dnKO are three other mouse models which depend on the inhibition of the TGFβ pathway (163). Knocking down the receptor expression or its signal transduction molecules results in the absence of the anti-inflammatory action of TGFβ in the colonic epithelium in these mice. These models require a trigger from DSS or OZ to initiate inflammation. The colonic epithelium of these models is characterised by profound aberrant crypt architecture and dysplastic change with longstanding inflammation (164). The relationship between inflammation and development of neoplasia in this model further strengthens the hypothesis behind the link between SMAD proteins and UC related neoplasia. Adoptive T cell transfer method utilises the transfer of naïve T4 cells into the colons of immune supressed mice. This method is widely used due to the higher level of control over immunogenicity and the ability to precisely initiate the disease onset and carry out a well-characterized experimental time course.

Azoxymethane (AOM, methyl-mrthylimino-oxido-azanium), 1,2-dimethylhydrazine (DMH), and/or methyl azoxy methane (MAM) acetate are all chemicals used in inducing carcinogenesis in animal models. In murine models with induced colitis, azoxymethane (AOM) is the most commonly used gent as a pro-carcinogenic agent to induce CAC. AOM is an alkylating agent when given in conjunction with DSSS provides a model to study the development of inflammation induced CRC (165, 166). Although the CAC development in IL-10-KO mice with AOM and DSS are similar to those in humans with UC, a major distinction is the lack of mutated k-RAS, ACP, P53 and MSH genes. Instead the murine models have demonstrated activation of NF-κB and JAK/STAT3 pathways (167-169). Also mutations in the β-Catenin gene have

been demonstrated in AOM/DSS induced murine CAC (170, 171). Although β-Catenin is also involved in the principal pathway in sporadic CRC it is due to a mutation of the APC gene. In AOM/ DSS murine models it is caused by a mutation in the β-Catenin gene. NF-κB, JAK/STAT and β-Catenin all link inflammation to CRC indicating the effect of longstanding inflammation to CAC (172, 173). Other models such as IL-2, P53, $G_{\alpha i2}$ and APC knocked out murine models have also been used to study CAC. In APC-*Min* mice models, bowel tumours with chromosomal instability, microsatellite instability and nonsense mutations have been observed with induced inflammation. However the APC-*Min* mice develop majority of their tumours in the small bowel as opposed to colonic tumours in humans with APC mutation (166).

However, from over 50 available animal models none can precisely replicate the complicated immune reaction seen in human UC. All models are designed to target one particular immune pathway while the human form elicits effect on multiple fronts. Also, the life cycle of mice is limited to around 3 months, which prevents the studying of long-term effects of inflammation on the colon. The most accurate model to study UC *in vitro* would be an organoids model developed from epithelial cells of patients with UC, as it could simulate the precise behaviour of the human epithelial cells with space for manipulation of immune reactions. Also, a 3D organoids model could assist in studying the effect of the stroma on the immune reaction in UC (174). Initial studies on the feasibility of an organoids model for IBD has been successful although it is not yet in wide use. The application of novel genomic techniques such as CRISPR on the organoids models could enable the study of the effect of individual genetic alterations on the disease process and progression.

### 1.10 Epigenetic changes beyond DNA methylation in ulcerative colitis

Although DNA methylation is the most studied area of epigenetics, other changes such as histone modification and chromosomal rearrangement have been implicated in the disease initiation and progression. Methylation and acetylation of histones have shown to regulate gene expression that influences the differentiation of intestinal epithelium (175). The polycomb repressive complex 2 (PRC2) causes tri-methylation of histones leading to suppression of terminal differentiation of intestinal crypt cells. It has been demonstrated that in IBD, Enhancer of Zeste Homolog2 (EZH2) which is the catalytic subunit of PRC2, is downregulated in humans, and the knockdown of the same in murine models have shown to induce a severe form of colitis with DSS (176). In the same study, the overexpression of EZH2 has shown the reverse with protection against severe colitis. Depletion of histone deacetylase complex (HDAC) 1 and 2 have shown to increase the severity of DSS induced colitis in mice (177). Chromosomal rearrangement influences the intestinal epithelial differentiation and maintenance of its barrier function (178). The depletion of chromosome associated protein (CAP) D3, a subunit of condensin, a protein involved in chromatin organisation, was observed in human colonic epithelium with active UC, which led to the postulation of a link between condensin and pathogenesis of UC (179). CAP-D3 is also found to repress intracellular proteins that block the autophagy function and is found to be important in suppressing bacterial infections such as of *E. Coli* in the gut. The function of both effector and regulatory T4 cells is also affected by histone modification and chromosomal rearrangement resulting in implications in UC. Depletion of the Jumonji Domain-containing protein 3 (JMJD3), a histone demethylase that regulates the differentiation of Th1, Th2, Th17 and regulatory T cells (Treg), results in a milder colitis in mice (180). Although the effect of JMJD3 is

seen on both pro- and anti-inflammatory T cells, the overall net effect of this histone modification appears to be anti- inflammatory by suppressing pro-inflammatory T cells. The histone lysine methyltransferase G9A causes histone methylation at the promoters of several genes including Fox3P, which inhibits the TGFβ1, mediated differentiation of Th17 and Treg cells. Loss of G9A function in T cells enhances Th17 and Treg differentiation that attenuates colitis. This mechanism has been proposed as a therapeutic target for IBD after demonstrating that pharmacological inhibition of G9A caused a reduced inflammatory response in murine models (181).

Imbalance between pro- and anti-inflammatory cytokines results in intestinal inflammation. IL6, a pro-inflammatory cytokine, can be inhibited by methylation of histones in the promoter region of the coding gene with the use of Atractylodin. Murine models show a reduced intestinal inflammation with inhibition of IL6 via histone methylation in TNF-α induced fibroblasts (182). Also, the Zinc finger protein KLF10 is important in inducing acetylation of histones, which promotes the transcription of TGFβRII receptors that maintain the anti-inflammatory action of TGFβ1 in macrophages. Depletion of KLF10 in macrophages results in an enhanced DSS-induced inflammation (183).

Early evidence exists to suggest the implications of chromatin rearrangement in the pathogenesis of IBD. Regulatory elements found in the noncoding regions of chromatin being differentially arranged leading to varied gene regulation is implicated in the clinical variability seen in IBD (184, 185).

# 2. Aims

## 2. Aims

Many investigators have searched for methylation markers that differentiate dysplasia and cancer from other tissues. While these markers while help to identify the pathophysiological process, they will not be helpful in predicting disease progression. To predict the progression, it is important to identify differentially methylated genes that will be present from the inflamed stage of the disease. Also, these markers should not be age related methylation changes and should be absent in normal colonic mucosa. In an inflammatory process, many genes can be differentially methylated as a cause for or as a result of the inflammation. Therefore, it is pertinent to recognise genes that are related to carcinogenic process based on existing evidence.

Hence in this project I aim to undertake a pilot study to identify the most potential candidates, which will need to be validated in a larger cohort of patients.

Specifically, I aim to perform whole genome bisulphite sequence on DNA extracted from inflamed, dysplastic and cancer tissue from UC affected patients along with control samples (epithelial cells and blood) from the same patients to exclude age related methylation changes. Furthermore, colonic epithelium from a ‘normal’ control (individual without UC) along with blood from the same patient as a control to exclude age related methylation changes will be sequenced. I intend to identify differentially methylated genes that are common to all three stages of the CAC process and not present in normal tissue. These can be regarded as early UC related methylation changes leading to CAC that have a predictive value.

A second aim of this study is the evaluation of the role of SMAD7 in CAC, as this molecule has a potential to influence DNA methylation via TGFβ pathway has

generated clinical interest. However, the inhibition of SMAD7 has been proposed as a therapeutic strategy without much inquiry into its effect on carcinogenesis in UC.

Based on the current literature on the role of SMAD7 in carcinogenesis, I hypothesise that the expression of SMAD7, known to increase during inflammation, may vary along deferent stages of CAC and possibly have an effect on the gene methylation in the colon as well as shown in breast cancer. SMAD7 has been shown to mediate the differential methylation of epithelial mesenchymal transmission (EMT) related genes CDH1, CGN and CLDN4. SMAD7 as shown in breast cancer and sporadic CRC may protect against the development of CAC through influencing the methylation of cancer related genes. Inhibiting its action without further studying this phenomenon would result in enhanced neoplastic transformation of the UC epithelium. Therefore I aim to study the expression of SMAD7 protein in multiple colonic samples across different stages of CAC using immunohistochemistry. Furthermore, I will be correlating the SMAD7 mRNA levels with the intracellular protein expression. Finally, As the third aim, I will determine the methylation status of the abovementioned EMT-related target genes in the same bisulphite whole genome sequencing dataset analysed above.

# 3. Materials and methods

## 3.1 Patient recruitment for fresh tissue collection

Ethical clearance for patient recruitment for fresh tissue collection was obtained from the Health Research Authority (HRA). A research proposal was submitted through the Imperial College NHS trust research and development unit, which was assessed by the West Midlands Black Country ethics review committee and approval was granted (17/WM/0206). Subsequently the HRA (IRAS – 219482) and institutional approval was granted for the project to go ahead.

Patients with UC who were scheduled to undergo colectomy for uncontrolled inflammation or dysplastic changes as well as those undergoing colectomy for non-UC related, benign conditions (chronic constipation, diverticular disease) were approached at the clinic and interest to participate was sought. To those who expressed interest a patient information sheet was handed over by the investigator. Those who consented to take part were approached again on the day of surgery and explained the procedure. Consenting patients signed a form allowing me to acquire a 5ml sample of blood and two sections of 2 X 2 cm size of mucosa from the colon, once the specimen was delivered.

## 3.2 Processing of fresh tissues specimens

### 3.2.1 Transportation of fresh frozen samples

Tissue sections were frozen using either isopentane (2-Methylbutane) or optimal cutting temperature (OCT) embedding compound depending on time interval between specimen acquisition and delivery. Specimens, which could be delivered to the laboratory within an hour of acquisition, were delivered in 0.9% NaCl (saline) solution and frozen using isopentane (2-Methylbutane). Those that had to be delayed

due to logistical reasons were snap frozen using optimal cutting temperature (OCT) at the operating theatre.

### 3.2.2 Freezing of samples using isopentane (2-Methylbutane)

From the surgical specimen a section of 2 X 2 cm size of mucosa was excised and delivered in 0.9% saline on ice. The specimen was dissected in the primary tissue culture cabinet to obtain only the mucosa. Isopentane in a Falcon tube was immersed in liquid Nitrogen till it was crystalised. The specimen was then immersed in crystallising isopentane and the container re-submerged in liquid nitrogen. When the content was completely frozen the cube was taken out into the safety hood and allowed to evaporate leaving the frozen tissue specimen. The frozen sample was labelled and stored in -80$^{o}$C freezer storage.

### 3.2.3 Freezing of samples using optimal cutting temperature (OCT) solution

A cryo-mould was filled with OCT and the tissue specimen was orientated inside the mould to be completely covered by the liquid. The mould was immediately placed on dry ice and allowed to freeze. The frozen block of tissue was then marked and stored in -80$^{o}$ C freezer storage.

All specimens were serially sectioned to 16 µm on MembraneSlides (Zeiss) for laser dissection and to 4µM sections on glass slides for staining.

## 3.3 Micro dissection of epithelial cells

Epithelial cells were micro dissected with laser capture microdissection (LCM), prior to DNA extraction in order to avoid the contamination with stromal tissue.

### 3.3.1 Staining of fresh tissue sections for cytochrome C oxidase

The fresh tissue sections were stained for the mitochondrial enzyme cytochrome C oxidase (CCO) to identify the epithelial layer for laser capture. The high concentration of the mitochondria in the active epithelial layer of colon tissue is exploited in this staining method. As the tissue is frozen fresh the enzyme activity is preserved enabling this method. This staining does not cause structural changes in the cells, which is crucial as they are used for genomic analysis. The oxidation of 3,3'-diaminobenzidine (DAB) by CCO is detected as brown staining of the epithelial layer (186). The frozen slides were defrosted and air-dried. 1.2 mL of cytochrome solution was mixed with 4 mL of DAB solution and a small amount of bovine catalase enzyme was added. The preparation was filtered using a 0.45 µm filter and applied on the tissue sections followed by incubation for 15 minutes in a humidified chamber at 37$^{o}$C. The solution was washed off with Phosphate buffered saline (PBS) and the slides air dried before storing at -80$^{o}$C.

### 3.3.2 Laser capturing of epithelial cells

Zeiss Palm® Laser capture microscope (Zeiss, Munich, Germany) was used for dissection. Dark brown staining from CCO allowed clear definition of the epithelium (**Figure 5**). The epithelial layer from the specimens was marked manually at 10X magnification. The areas of interest were dissected using pre-optimised energy level and catapulted in to S500 collection tubes with adhesive caps (Zeiss, #415190-9211-000). The robot programmable logic controllers (PLC) mode was used to catapult the

cut sections automatically.

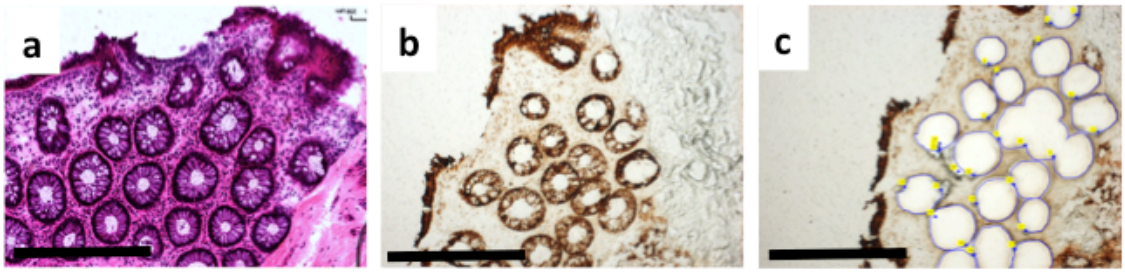



**Figure 5. Laser capture micro dissection of colonic epithelial cells**

Serial sections of, a) H and E staining of a fresh colonic epithelial tissue section. b) CCO stained serial section of normal colonic epithelium (Scale bar = 250µm). c) The epithelial cells of the individual crypts are laser dissected and captured. (scale bar – 250 µm). The CCO staining in b) and c) clearly demonstrating the epithelial cells in dark brown.

### 3.4 Processing FFPE tissue

FFPE tissue blocks were acquired from the pathological archives of St Mark's hospital through institutional ethical approval (RD16/084) and material transfer agreement. All specimens were sectioned to 16µm serial sections on MembraneSlides (Zeiss) for laser dissection and to 4µm sections on normal glass slides for staining.

#### 3.4.1 Quality assessment of genomic material of FFPE sections for sequencing

To test the quality of the genomic material in the sections, a multiplex polymerised chain reaction (PCR) for 100, 300 and 400 bp segments of the Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene were performed. One FFPE tissue section per sample was used to perform DNA extraction using the QIAamp DNAeasy micro kit (#56304, Qiagen, Hilden).

The tissue sections were first deparaffinised by incubating in xylene followed by 100% alcohol. The sections were then scraped using a scalpel blade and the DNA was extracted using the Qiagen kit as specified by the manufacturer. The tissue sampled by LCM was incubated with proteinase K DNA extraction buffer at 56°C overnight. After neutralisation of proteinase K, 1µl of carrier RNA (Qiagen) was added to the lysis buffer to enhance binding of the DNA to the column membrane. Then, two centrifugation steps were used to clean the DNA bound to the column membrane using two distinct wash buffers to enhance the purity of the eluted DNA. Finally, in 20µl of RNase-free H20 (Sigma-Aldrich), DNA was eluted straight to the membrane after an hour of incubation. Custom made primer sequences were acquired for overlapping segments of GAPDH gene (chr12) based on a previously published data (187). The primer sequences are listed in table 2.

| | Sequence |
|---|---|
| 100 F | GTTCCAATATGATTCCACCC |
| 100 R | CTCCTGGAAGATGGTGATGG |
| 300 F | AGGTGAGACATTCTTGCTGG |
| 300 R | TCCACTAACCAGTCAGCGTC |
| 400 F | ACAGTCCATGCCATCACTGC |
| 400 R | GCTTGACAAAGTGGTCGTTG |

**Table 2. Sequences of the primers used for multiplex PCR.**

The primer sequences were custom made based on previously published research (187). The sequences are for three non-overlapping segments of the GAPDH gene that is ubiquitously present in the cells.

Multiplex PCR was performed on 200 ng of purified DNA with a primer concentration of 0.25 µM in a mix of 2.5 µl of 10X PCR buffer (New England Biolab), 0.5 µl of dNTP mix (New England Biolab), 0.5 µl of each primer, 0.2 µl of Taq Polymerase (New England Biolab) and a final volume of 25 µl. The PCR reaction was carried out in the Tetrad 2 thermal cycler (Marshall scientific) under the following conditions: Denaturation was performed at 94$^{o}$ C for 4 minutes, denaturation at 94$^{o}$ C for 1 minute, annealing at 56$^{o}$ C for 1 minute and elongation at 72$^{o}$ C for 3 minutes repeated for 34 cycles. Termination step was performed at 72$^{o}$ C for 7 minutes followed by ending the cycle at 10$^{o}$ C.

The product was separated by gel electrophoresis using a 1.5% agarose gel in Tris/Borate/EDTA (TBE), after addition of 6 µl of bromophenol blue (Sigma) loading dye (5X) 10 µl of sample was electrophoresed. A 100bp DNA ladder (Sigma) and samples were electrophoresed at 100mV. The gel was visualised under ultraviolet (UV) image capture device (**Figure 6**).

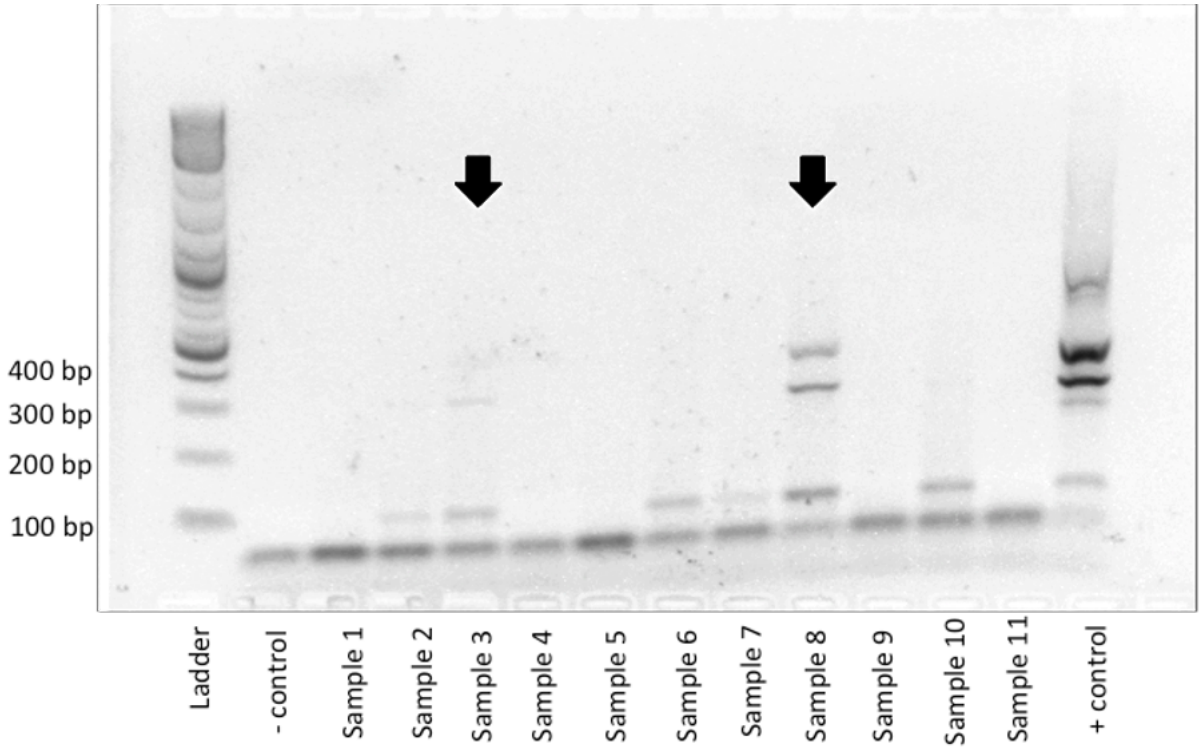



**Figure 6. Multiplex PCR to assess the quality of DNA in FFPE segments for NGS**

Specimens with at least 300bp segments of GAPDH were chosen for NGS (White arrows indicating the specimens with 300 and 400 bp segments).

### 3.4.2 Methyl green staining of FFPE sections

The FFPE sections were stained with methyl green to visualise the epithelial component for laser capture. Since the FFPE tissues are fixed the CCO enzyme activity cannot be utilised for staining. Methyl green has a specific affinity to polymerised DNA without causing structural damage to genomic material. It also binds with less affinity to protein structural components of the tissue providing a clear outline of the tissue structure with enhancement of the nuclear material(188). The sections were deparaffinised by submerging in xylene and serial dilutions of alcohol. Diluted (1:1) Methyl Green (Sigma-Aldrich) was applied to the sections and left for 3

minutes. The sections were dehydrated using serial dilutions of alcohol (70%, 90%v and 100%) and air dried before storage in +4$^{o}$C.

### 3.4.3 Laser capturing of epithelial cells

Zeiss Palm® Laser capture microscope (Zeiss, Munich, Germany) was used for dissection. Green staining from methyl green allowed clear definition of the epithelium (**Figure 7**). The epithelial layer from the specimens was marked manually at 10X magnification. The areas of interest were dissected using pre-optimised energy level and catapulted in to S500 collection tubes with adhesive caps (Zeiss, #415190-9211-000). The robot programmable logic controllers (PLC) mode was used to catapult the cut sections automatically.

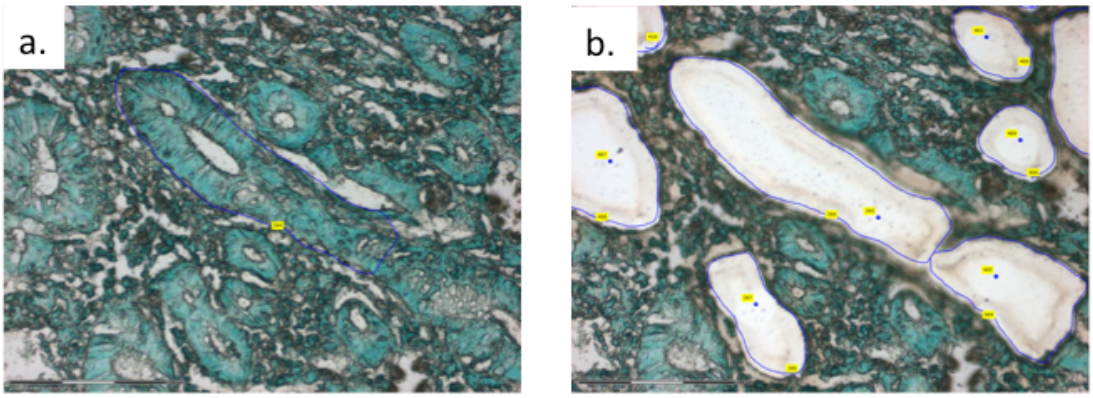



**Figure 7. Laser capture microdissection of FFPE sections**

Laser capturing of epithelial cells from a) an FFPE section stained with methyl green. b) Shows the selected epithelial cell sections being cut and extracted leaving the stroma (Scale bar 250 µm).

## 3.5 Immunohistochemistry (IHC) staining

### 3.5.1 Optimisation of antibodies

Optimisation of the anti SMAD7 and pSMAD3 antibodies was performed on hepatocellular carcinoma and human tonsillar tissue kindly gifted by George Elia, Barts Cancer Institute. Human liver cancer tissue has a consistent high expression of the protein SMAD7 and has been used previously. Similarly, human tonsillar tissue expresses pSMAD3 protein at high concentrations. Anti SMAD7 (Monoclonal Mouse IgG2B, #MAB2029, R&D systems, Minneapolis, MN) and pSMAD3 (Rabbit monoclonal [EP823Y] to Smad3 [phospho S423 + S425], Abcam, Cambridge, MA) were serially diluted (0.008 - 0.33 µg/µl and 1:25 - 1:100) in 10% goat serum and incubated at 4C overnight.

Optimal staining intensity was decided by the specificity of the cellular staining and lack of background staining. Negative control staining was performed with Isotype Control for mouse (Mouse IgG1 kappa [MOPC-21] , ab106163, Abcam, Cambridge, MA) and rabbit antigens (Recombinant Rabbit IgG, monoclonal [EPR25A] - ab172730, Abcam, Cambridge, MA). Working concentration of 0.33 µg/µl was chosen for anti-SMAD7antibody and 1:50 dilution was chosen for anti pSMAD3.

### 3.5.2 Immunohistochemistry staining of FFPE sections

Previous studies by Monteleone and colleagues have assessed the expression of SMAD7 in mucosal samples from patients with IBD using 8 samples for test and control groups, where all samples demonstrated a significantly higher values compared to non-inflammed colons (142). The number of slides required for the current study was calculated based on a power of 80% and α of 0.05, taking into

consideration the mean values of the IHC staining intensity of 30% as the cut-off for the negative staining versus 50% for the moderately positive. A difference in the mean of 20% with a standard deviation of 10% the required sample size is 4 per group according to the Pocock equation owing to the significant difference amongst the groups (239).

$$n = f(\alpha, \beta)\chi \frac{2\sigma^2}{(\mu 1 - \mu 2)^2}$$

** f(α,β)= factor based on error and power [ 7.85 for α=0.05 and β=80%] , (μ1 – μ2)= expected difference in means, σ = standard deviation.*

I have included a total of 53 samples from non-inflamed (n=12), inflamed (n=12), dysplastic (n=12) and colitis associated cancer specimens (n=17). Sections of 4 μm thickness were deparaffinised in xylene (Sigma-Aldrich, Dorset) and rehydrated in decreasing concentrations of ethanol (100%, 90% and 70%). After washing antigen retrieval was done by incubating the sections in boiling sodium citrate buffer, pH 6.0 (2.57g of sodium citrate dissolved in 1L of distilled water and pH corrected to 6.0 by adding 0.1 M HCl) in a microwave at 800 kW for 20 minutes. After washing in tap water followed by two 5 minutes washes in PBS sections were blocked in 10% goat serum (Sigma-Aldrich) diluted in PBS for 2 hours at room temperature (RT). Incubation with the primary antibody (**Table 3**) was performed overnight at 4$^{o}$C. Following washes in PBS with 0.5% Tween 20 (Sigma-Aldrich) to remove unbound antibodies and remove possible unspecific reactions, block of endogenous peroxidise was performed with 3% hydrogen peroxide ($H_2O_{2)}$ in methanol for 15 minutes at RT. Secondary antibody was applied diluted in 10% goat serum (**Table 3**). Sections were then incubated in streptavidin conjugated HRP (Invitrogen, Carlsbad, CA) HRP (1:200) for 40 minutes at RT and developed with 3,3'-Diaminobenzidine

(DAB, Sigma-Aldrich),) for 10 minutes at RT. Counterstaining was performed with Gill's haematoxylin (Sigma-Aldrich) for 3 minutes followed by mounting of a coverslip with quick-hardening mounting medium (Eukitt, Sigma-Aldrich).

| Antibody | | Dilution | Supplier | Catalogue number |
|---|---|---|---|---|
| Anti-SMAD7 | Mouse monoclonal | 0.33 µg/µl | R &D, Minneapolis, MN | MAB2029 |
| Anti- pSMAD3 | Rabbit polyclonal | 1:50 | Abcam, Cambridge, UK | Ab52903 |
| Anti-mouse antibody | Goat, anti-mouse IgG whole molecule | 1:500 | Sigma-Aldrich | M5899 |
| Anti-rabbit antibody | Goat- anti rabbit | 1:1000 | Abcam | Ab97049 |

**Table 3. Antibodies used for immunohistochemistry (IHC) of colon sections**

### 3.5.3 Quantification of staining intensity and protein expression

Stained slides were scanned with NanoZoomer (Hamamatsu Photonics, Shizuoka, Japan) and saved. Visiopharm (6.9.0.2808, Hoersholm, Denmark) image analysis software was used to quantify the staining intensity of the sections, representing the expression of each protein. In-built standard applications are available for the nuclear and cytoplasmic assessment. Application was tuned to identify the DAB staining of the cytoplasm and nucleus separately using smaller regions. The epithelial cells were selected as a region of interest (ROI), and the tuned applications were used to assess the stained regions. The cytoplasmic and nuclear regions stained positively and negatively were highlighted in separate colours **(Figure 8)**. An initial set of 30 slides was validated independently by a histopathologist trained in gastrointestinal

pathology for accuracy of machine learning. Visual analysis of the stained slides under 10X magnification of histopathological sections, which is the current standard in clinical practice, was considered the reference. The visual assessment of stained slides as done in clinical practice was undertaken with arbitrary values of 0, + and ++ based on the level of staining. Staining of slides <30% is graded as 0, between 30 – 70% as + and > 70% as ++. The slides were independently analysed in three separate rounds with the image analyser for the percentage of staining and manually for visual assessment. After comparing the two methods the image analyser settings was fine-tuned to match the visual analysis. The staining intensity was expressed as a percentage of stained cells over the total number of epithelial cells in the selected field.

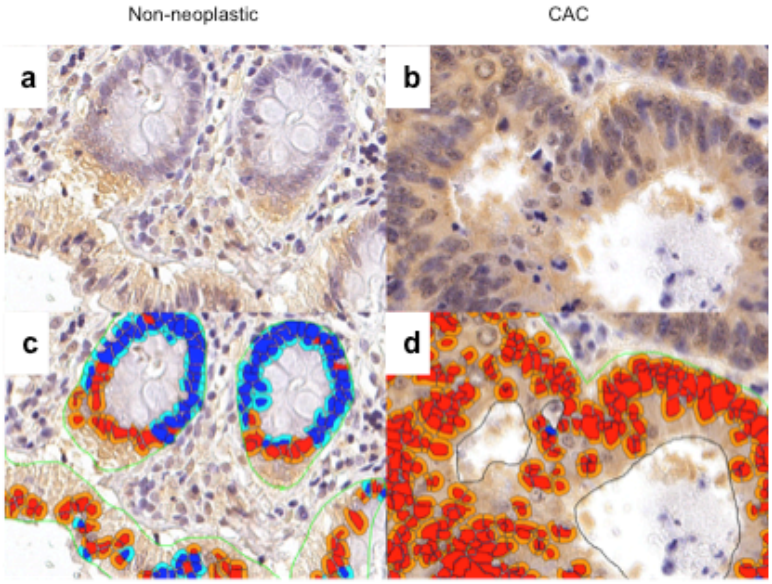



**Figure 8. Analysis of staining intensity using image analysis software**

The pictures show a representative example of SMAD7 staining and Visiopharm software detection. The stained cytoplasm is marked in in orange while the nuclei of the stained cells appear in red. The unstained cytoplasm and nuclei are highlighted in light blue and dark blue respectively. Faintly stained cells in the non-inflamed/ non-

neoplastic colonic section (a) and the intensely stained cells in CAC (b) accurately detected by the software (c, d).

Expression of SMAD7 was assessed in the epithelial cells by quantifying the cytoplasmic staining, as it is a regulatory protein that has its function in the cytoplasmic level in the TGF pathway. pSMAD3 levels were assessed using nuclear staining as the phosphorylated form of SMAD3, when not inhibited by SMAD7, migrates to the nucleus.

## 3.6 *In situ* hybridisation (ISH)

### 3.6.1 Workflow optimisation for RNAscope

Workflow optimisation was carried out using fixed HeLa cells provided by the manufacturer. Slides were baked at 60$^{o}$C for one hour, deparaffinised in xylene and rehydrated in decreasing concentrations of ethanol. Antigen retrieval was performed using the antigen retrieval reagent and block of endogenous peroxidase performed for 10 minutes by applying $H_2O_2$ on the tissue sections. Sections were then incubated in protease A for 30 minutes in a humidifying chamber at 40$^{o}$C. Two separate slides were individually incubated with a positive control probe (PPIB) and a negative control probe for 2 hours at 40$^{o}$C. For each target gene, six signal amplification probes were applied consecutively for 30 and 15 minutes alternatively (probes 1-4 at 40$^{o}$C and probes 5-6 at room temperature). Signal was developed using the DAB A and B solutions provided. Counter staining was performed with Gill's haematoxylin solution for 3 minutes and serially dehydrating in increasing concentrations of ethyl alcohol.

### 3.6.2 Evaluation of SMAD7 mRNA expression in FFPE sections by ISH

Two serial sections of each sample were used as positive and negative control. Slides were baked at 60$^{o}$C for one hour and deparaffinised in xylene. Rehydration was carried out in reducing concentrations of ethanol (100%, 90% and 70%) and treated with hydrogen peroxide for 10 minutes. Antigen retrieval was carried out using the antigen retrieval reagent (Advanced Cell Diagnostic,) at 102$^{o}$C for 15 minutes. Protease A was applied for 30 minutes and incubated in a humidifying chamber at 40$^{o}$C. Positive control probe (*PPIB*), *SMAD7*gene specific oligo probe and the negative control probe were applied to the paired slides and incubated for 2 hours at 40$^{o}$ C. The positive control probe PPIB which is a ubiquitous housekeeping gene was used to select the sections with suitable for ISH and to normalise the target signal for time dependent degradation of mRNA. Sections that did not give a signal for the positive probe were excluded as they were of poor quality. Slides were washed in RNAScope wash buffer and the signal amplification probes were applied as per manufacturer specifications. Signal was detected with 3,3'-Diaminobenzidine (DAB) provided with the kit. Counter staining was performed with Gill's haematoxylin solution for 3 minutes and serially dehydrating in increasing concentrations of ethyl alcohol.

### 3.6.3 Quantification of ISH signal

SMAD7 mRNA is represented by brown single or clustered dots on the sections. One distinct dot indicates a single copy of the target mRNA. The signal was normalised to the signal of the positive control *(PPIB)*. Staining intensity quantification was performed using Image J software **(Figure 9)**. The epithelial cells of both control and test sample were first selected while the background was omitted. The values for

HBS (H- hue, B- brightness and S-saturation) were adjusted so that only areas of signal (brown) were uniquely selected. The number of pixels in the signal for a single copy of mRNA was calculated estimating the size of a single dot, by averaging the number of pixels from 20 single dots (22 pixels). The number of mRNA copies in the field was calculating at by dividing the total positive areas selected by 22. The number of epithelial cells in the region of interest was counted using the cell counter plug-in and thereby the number of signals per cell was calculated. Finally, the ratio between signals per cell of PPIB and signals per cell for SMAD7 mRNA was calculated.

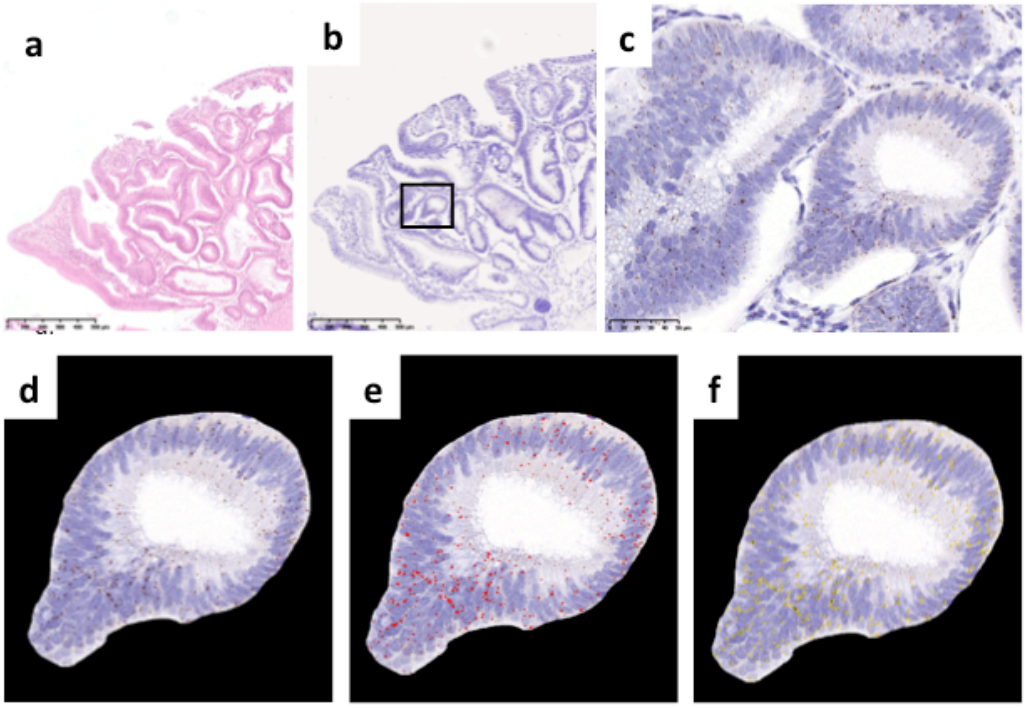



**Figure 9. Quantification of *In situ* hybridisation (ISH) using the ImageJ software**

a) H&E staining of a dysplastic lesion in a patient with UC. (b) ISH of a serial section from the same lesion (selected area in-set) and (c) magnified (40X) image of the selected area. (d) The epithelium is selected as region of interest. (e) The brown stained areas are detected and marked with red. (f) the detected surface area is demarcated with a yellow border for the calculation of pixels.

## 3.7 Processing of blood samples

### 3.7.1 Separation of the buffy coat

Blood samples were transported at +4$^{o}$ C to the laboratory. Sample was centrifuged at 2500rpm for 10 minutes at room temperature. The sample separates into red cells at the bottom, buffy coat containing nucleated white cells above it as a hazy layer and clear plasma as a supernatant. Under the primary tissue culture hood the plasma and buffy coat were separated by pipetting and stored in -80$^{o}$ C freezer.

## 3.8 Sample processing and whole genome bisulphite sequencing

Extracted DNA was required to be quantified and run through library preparation process prior to sequencing with Illumina 4000 platform. The library preparations had to be subjected to quality check prior to sequencing to ensure high quality output of sequencing data.

### 3.8.1 DNA extraction from micro dissected epithelial cells from fresh tissue sections

DNA was extracted with a column-based method using the QIAmp DNA Micro kit (Qiagen, Hilden, Germany), as specified by the manufacturer. The tissue sampled by LCM was incubated with proteinase K DNA extraction buffer at 56°C for 3 hours. After neutralisation of proteinase K, 1µl of carrier RNA (Qiagen) was added to the lysis buffer to enhance binding of the DNA to the column membrane. Then, two centrifugation steps were used to clean the DNA bound to the column membrane

using two distinct wash buffers to enhance the purity of the eluted DNA. Finally, DNA was eluted in 20µl of RNase-free H20 (Sigma-Aldrich), after an hour of incubation.

### 3.8.2 DNA extraction from micro dissected epithelial cells from FFPE tissue sections

QIAamp DNAeasy micro kit (Qiagen) was used according to manufacturer's recommendations. Briefly, collected laser captured cells were lysed in lysis buffer containing proteinase K at 56$^{o}$C overnight. Proteinase K was then denatured at 90$^{o}$C for 1 hour, and the DNA was obtained by binding to a column and subsequently washed in a series of provided wash buffers. For genomic DNA elution, 20 µl of DNase/RNase free water was applied directly to the membrane of the QIAamp MinElute column and left at room temperature for 1 min. The column was centrifuged at 12000 rpm for 1 minute and the flow through collected to a sterile 1.5 mL microcentrifuge tube was stored at -20$^{o}$C.

### 3.8.3 DNA extraction from buffy coat

QIAamp DNAeasy micro kit (Qiagen, Hilden) was used. Frozen buffy coat was defrosted to room temperature and a 10 µl aliquots was mixed with 90 µl if buffer ATL in a clean micro centrifuge tube. 10 µl of proteinase K and 100 µl of buffer AL were added and mixed by vortexing. The tubes with the mix were sealed with paraffin membranes and incubated at 56$^{o}$C for 10 minutes and briefly centrifuged. The DNA was obtained by binding to a column and subsequently washed in a series of provided wash buffers. For genomic DNA elution, 20 µl of DNase/RNase free water was applied directly to the membrane of the QIAamp MinElute column and left at room temperature for 1 min. The column was centrifuged at 12000 rpm for 1 minute

and the flow through collected to a sterile 1.5 mL microcentrifuge tube was stored at -20$^{o}$C.

### 3.8.4 Quantification of DNA

Qubit 4® Fluorometer (Invitrogen, Carlsbad, CA) with the Qubit® dsDNA HS (High Sensitivity) assay kit (Thermo scientific) was used to measure the double stranded DNA concentration. Working solution was prepared by diluting the Qubit® dsDNA HS Reagent 1:200 in Qubit® dsDNA HS Buffer. Each tube for standards were added 190 µl of the working buffer solution and 10µl of Standards 1 and 2. Each sample tube was added 199µl of the buffer solution and 1µl of sample. Tubes were pulse vortexed for 15 seconds. The tubes were left for 2-3 minutes in room temperature. The high sensitivity dsDNA function was selected, and the samples were measured following calibration with the standards. Concentrations were recorded as ng/µl.

### 3.8.5 Library preparation for WGBS

Library preparation for next generation sequencing using Illumina HiSeq4000 platform was performed with Pico Methyl-Seq Library Prep Kit (Zymo, Irvine, CA, D5455 & D5456). The kit is specific for small amounts (10 pg to 100 ng) of fragmented genomic DNA from FFPE sections. Library preparation process included: fragmentation, bisulphite conversion, addition of adaptors and addition of index sequences.

Starting amount of 100ng of genomic DNA concentrated to a volume of 20 µl was used for all samples. The genomic DNA was mixed with 130 µl of lightning conversion agent in a PCR tube and incubated in a thermo cycler for 8 minutes at 98$^{o}$ C and 1 hour at 54$^{o}$ C. The fragmented and bisulphite converted sample was

added in to 600 µl M-Binding Buffer in a Zymo-Spin® IC Column in a Collection tube and mixed thoroughly by inverting the tube. Following a washing step, 200 µl of L-Desulphonation Buffer was added to the column and incubated at room temperature for 15 to 20 minutes. Washing step was repeated and library samples were eluted with 8 µl Elution Buffer.

***Amplification with the PrepAmp Primer***

The flow through containing the bisulphite converted DNA was mixed with 2 µl of PrepAmp Buffer (5X) and 1 µl of PrepAmp Primer (40 µM). The priming mix was incubated in the thermo cycler for 2 cycles at temperature ramping at $0.1^{o}$ $CS^{-1}$with the lid temperature constant at $25^{o}$ C (Table).

***Purification with the DNA Clean-up & Concentrator (DCC)***

Product from the PrepAmp Polymerase Reaction was mixed with DNA Binding Buffer at a 1:7 ration. The mix was centrifuged in a Zymo-Spin IC column at 10000 g for 30 seconds and the column was washed with 200µl of DNA Wash buffer. The spin column was inserted in to a 1.5 mL tube and 12 µl DNA Elution Buffer was added directly to the membrane. After incubating for 1 minute the column was centrifuged for 30 seconds and the purified product was collected.

***Amplification***

The purified product was mixed with 12.5 µl of LibraryAmp Master Mix (2X) and 1 µl of LibraryAmp Primers (10 µM). The mix was incubated in a thermo cycler for 6 cycles to amplify. The amplified product was purified as described above using the DCC with 12 µl of elution buffer.

***Amplification with the Index Primer***

The purified product was then mixed with 12.5 µl of LibraryAmp Master Mix (2X) and 0.5 µl of Index Primer (10 µM). The kit comprises of 6 index primers named A-F. The mix was incubated for 9 cycles in a thermo cycler (Table) at variable temperatures. The PCR product was purified as described above using the DCC© with 12µl of elution buffer.

### 3.8.6 Quality assessment of library preparations

Agilent 2200 TapeStation (Agilent Technologies, Germany) with D1000 ScreenTape and D1000 Reagents were used. An aliquot of 1 µl of each sample was mixed with 3 µl of sample buffer along with a DNA ladder with a range of 35 – 1000 bp. After loading the screen tape and tips the samples were loaded and the machine programmed for reading. The quality report and fragment length graph for reach sample was extracted.

## 3.9 Whole genome bisulfite sequencing

WGBS was performed using the Illumina HiSeq4000 platform. Standard mode, 2 X 75 bp read lengths at 10X with 1.5 lanes per sample was assigned. The PhiX control was set at 20% as recommended by the manufacturer. The system performs 540 - 624 million reads per lane with 40 - 47 Giga bases output per lane. Sequencing was performed at the Imperial BRC genomic facility (189).

## 3.10 Bioinformatics

Beijing Genomic Institute (Shenzhen, Hong Kong) performed the initial bioinformatics analysis of producing clean methylation data. Pre-processing and analysis of data followed the Broad Institutes “GATK Best practice guidelines” for Next generation Sequencing. FASTQC software was used to validate the FASTQ data and filtering of raw data. The annotated human reference genome (GRCh37/hg19) was used for alignment of the clean data. The bioinformatics analysis pathway of the NGS data included quality check, filtering, clean up, alignment, cytosine information, depth of coverage and methylation level analysis **(Figure 10)**. BGI carried out the data cleaning, read alignments and identification of differentially methylated regions for each test sample obtaining methylation levels (M values). Analysing the variability in methylation between different stages, comparing the common elements and annotating the identified DMRs in to pathways were done by me using R open source statistical software.

After sequencing data was obtained, data filtering was performed, which removed the low-quality data. Then, the clean data was matched to a reference genome after ensuring the quality of the data was sufficient. Both sequencing quality and alignment quality was assessed before methylation analysis was carried out. Uniquely mapped data was used to derive the methylation information of the cytosine base (mC) through the whole genome. The mC data was used to analyse the differential methylation between samples and controls.

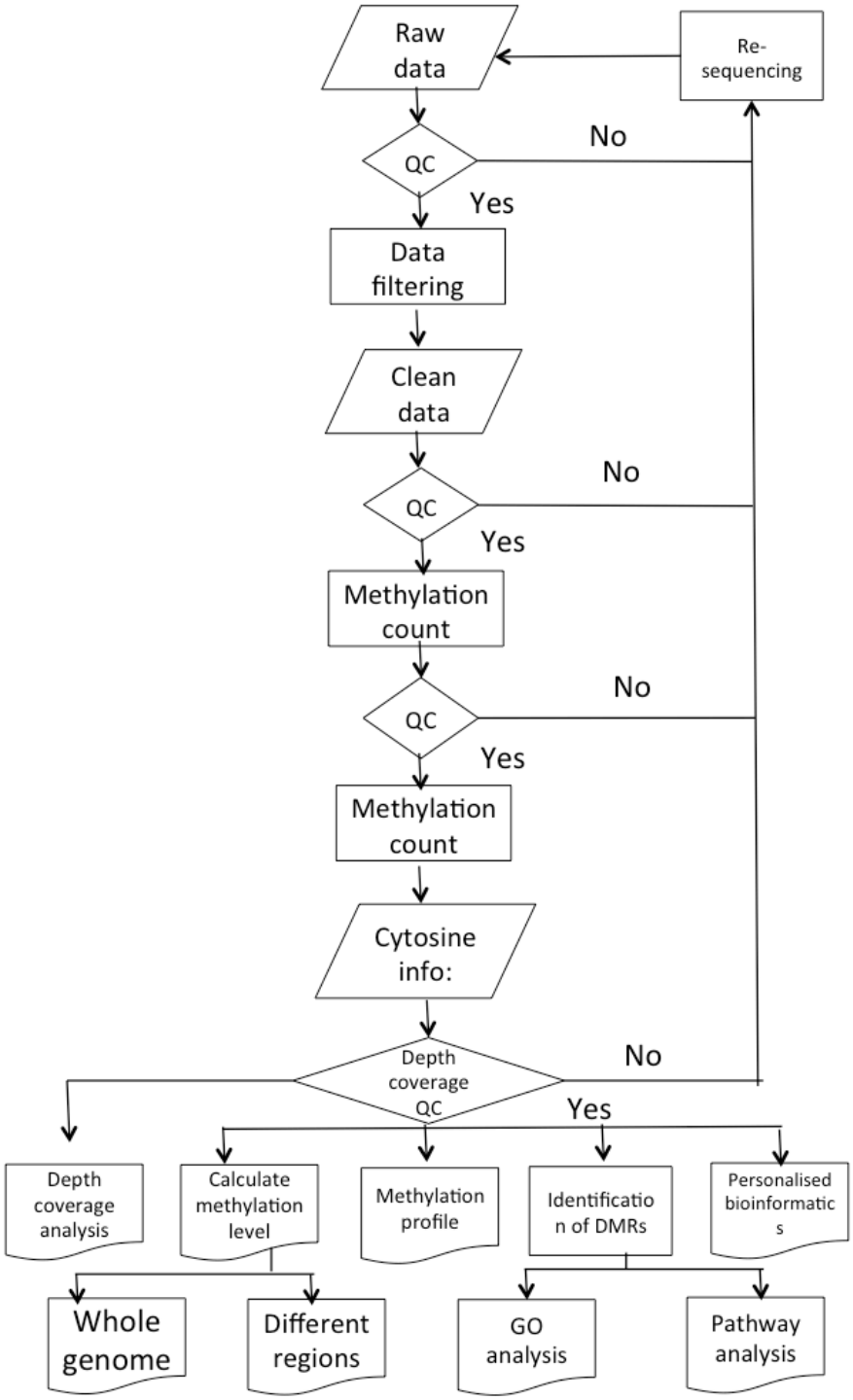



**Figure 10. Bioinformatics analysis pipeline of whole genome sequencing data**

Flow chart depicting the data analysis pathway. Each step of filtering, alignment, methylation count and depth coverage analysis were intervened by a quality checks (QC) step. Only the data that passed the QC were used to evaluate the methylation level and pathway analysis.

### 3.10.1 Data filtering

Data filtering includes removing adaptor sequences, contamination and low-quality reads from raw reads. Low quality reads include two types, and the reads, which met anyone of the two conditions below, were removed:

1) Unknown bases are more than 10%;

2) The ratio of bases whose quality was less than 20 was over 10%.

After filtering, the remaining "clean reads" were stored in FASTQ format.

### 3.10.2 Alignment

After filtering, the clean data was then mapped to the reference genome by BSMAP, and then remove the duplication reads and merge the mapping results according to each library. Here we calculate the Mapping rate and bisulphite Conversion rate of each sample.

**Reads Alignment process**

```
BSMAP parameters for PE reads: BSMAP -a filename_1.clean.fq.gz -b
filename_2.clean.fq.gz -o filename.sam -d
ref.fa
If use Hiseq2000, -z should be set 64;
samtools parameters:
samtools view -S -b -o filename.bam filename.sam
samtools sort -m 2000000000 filename.bam filename.sort
samtools index filename.sort.bam
```

Methylation level was determined by dividing the number of reads covering each mC by the total reads covering that cytosine (190), which was also equal to the mC/C ratio at each reference cytosine (191).

$$Rm_{average} = \frac{\mathrm{Nm_{all}}}{\mathrm{Nm_{all}+Nnm_{all}}} * 100\%$$

(***Nm*** represents the reads number of mC, while ***Nnm*** represents the reads number of non-methylated reads).

### 3.10.3 DMR detection

Differentially methylated regions (DMRs) are stretches of DNA in a sample's genome that have different DNA methylation patterns compared with other samples. DMRs are involved in genetic imprinting because they can be methylated in accordance with either the maternal or the paternal chromosome. The methylated allele is often, but not always, the silenced allele. Differences between the methylation pattern in the parental chromosome and the offspring's chromosome are considered epigenetic lesions. There has been some concern that artificial reproductive techniques increase the rate of abnormal methylation in DMRs, leading to an increased rate of disease (192, 193). A sliding-window approach was used to identify differentially methylated regions (DMRs) which contained at least five CG (CHH or CHG) sites and its methylation levels of two different samples were different (at least a twofold change). Fisher's exact test with a p value of <= 0.05 was used to assess the significance. In addition, both samples required to be non-hypomethylated in DMR discovery. Two nearby DMRs were considered interdependent and joined into one continuous DMR if the genomic region from the start of an upstream DMR to the end of a downstream DMR also had 2- fold Methylation level differences between sample and control (p value <=0.05). Otherwise, the two DMRs were viewed as independent.

After iteratively merging interdependent DMRs, the final dataset of DMRs were made up of those that were independent from each other.

### 3.10.4 Degree of difference in methylation level

The degree of difference of a methyl-cytosine (mCG, mCHG, mCHH) was calculated between the test and control samples while comparing Methylation level of DMRs in different samples by CIRCOS (circos.ca/software/download/circus).

$$\text{degree of difference} = \frac{\log_2 \text{Rm1}}{\log_2 \text{Rm2}}$$

(Rm1、Rm2 represent the Methylation level of methyl-cytosine for sample1 and sample2 respectively. Rm1(or Rm2) was taken to be 0.001 while it is = 0).

### 3.10.5 Gene ontology analysis

Gene Ontology (GO), which is an international standard gene functional classification system, offers a dynamic-updated controlled vocabulary, as well as a strictly defined concept to comprehensively describe properties of genes and their products in any organism. GO has three ontologies: molecular function, cellular component and biological process. The basic unit of GO is GO -term. Every GO -term belongs to a type of ontology. GO enrichment analysis provides all GO terms that significantly enriched in a list of DMR-related genes, comparing to a genome background, and

filter the DMR-related genes that correspond to specific biological functions. This method firstly maps all DMR related genes to GO terms in the database ( http://www.geneontology.org/), calculating gene numbers for every term, then uses hypergeometric test to find significantly enriched GO terms in the input list of DMR-related genes, based on ' GO ::TermFinder' (http://www.yeastgenome.org/help/analyze/go-term-finder), we have developed a strict algorithm to do the analysis, and the method used is described as follows:

Where N is the number of all genes with GO annotation; n is the number of DMR-related genes in N; M is the number of all genes that are annotated to certain GO terms; m is the number of DMR-related genes in M. The calculated p-value goes through Bonferroni Correction (194), taking corrected p-value ≤ 0.05 as a threshold. GO terms fulfilling this condition are defined as significantly enriched GO terms in DMR-related genes. This analysis is able to recognize the main biological functions that DMR-related genes exercise.

The annotation of the proposed candidate genes was done using the ToppGene Suit open source tool software. The target gene set was matched against the extended gene list from the Protein-protein interaction network (PPIN) as a test set. The p values for the gene enrichment analysis was accepted at 0.05 with the Bonferroni correction. The ToppGene Suit database is automatically updated and has been validated (195, 196).

### 3.10.6 Statistical analysis and graphs

The graphs were generated using ggplot2 package and maftools from Bioconductor in R (Verion 4.0.2, CRAN project, Vienna). The Kruskal-Wallis test was used to

compare the difference between the different tissue groups. The difference was regarded as significant if $P<0.05$.

## 3.11 Validation of methylation data with gene expression levels

To validate the observed methylation levels we carried out quantitative reverse transcriptase polymerised chain reaction (RT-PCR) for selected genes on the matched sample couples. Extracted RNA from fresh tissue was used for this purpose with custom designed primers.

### 3.11.1 RNA extraction

miRNeasy MiniRNA purification kit (#217004, Qiagen) was used. Trizol (Qiagen) 700 µl was directly added to the wells when harvesting the cells. Cells were scraped off the wells and the lysate collected in to a 1.5mL micro centrifuge tube. The lysate was mixed till it was homogenous and 140**µ**l of chloroform was added under a fume hood. After mixing well the tubes were centrifuged at 12,000 x g in 4°C for 15 minutes. The upper aqueous phase was transferred in to a new tube and 1.5 volumes of 100% ethanol were added. The sample was pipetted in to a miRNeasy Mini spin column and centrifuged at 8000 g for 30 seconds at room temperature. On-tube DNase digestion was performed by adding 10 µl of DNase I and 70 µl of RDD directly on to the membrane after a single wash with 350 µL of Buffer RWT. This step was followed by a single wash with Buffer RWT and two washes with Buffer RPE. RNA was extracted with 30µl RNase-free water.

### 3.11.2 Reverse transcription

RNA was retro transcribed using Maxima First Strand® High Capacity cDNA reverse transcription kit (#K1641, Applied Biosystems, Life Technologies, Carlsbad, CA, USA), as directed by the manufacturer. In this assay, 1µg total RNA from the sample per was used. A master mix containing 2 µl of 10x RT buffer, 1 µl of dNTP Mix, 2 µl of 10x RT random primers, 1 µl of MultiScribe reverse transcriptase and 1µl of RNase inhibitor was prepared per sample. Then, 1µg total RNA with RNase-free water was added to the master mix up to 20 µl per reaction.

### 3.11.3 **Quantitative real-time polymerase chain reaction**

The cDNA was amplified using the SYBR® Green Real-Time PCR Master mixes (#4309155, Life Technologies, USA) with the primers. Reaction was carried out in a 20 µl final volume using pretested primer sequences **(Table 4 and 5).**

| Reagents | Volume (µl) |
|---|---|
| Fast SYBR green master mix | 10 |
| 10 µM F primer (1 µM) | 1 |
| 10 µM R primer (1 µM) | 1 |
| Template cDNA (200ng) | 4 |
| Nuclease Free water | 4 |

**Table 4. Reagent mix for qPCR**

The thermo cycler conditions were set to 20 seconds at 95°C; 3sec at 95°C and 30sec at 63°C (40 cycles); 15 seconds at 95°C. Standard curve and melting curve were generated and analysed using StepOne Plus Real-Time PCR System software (Applied Biosystems, MA, USA).

| Gene | Primer sequence |
|---|---|
| TACSD2 - F | TATTACCTGGACGAGATTCCCC |
| TACSD2 - R | CCCCGACTTTCTCCGGTTG |
| eIF6 - F | CAATGTCACCACCTGCAATG |
| eIF6 - R | AGTCATTCACCACCATCCCA |
| LFNG - F | ACCTCACAACAGCAAACCCA |
| LFNG - R | GGCTTTCAACGACTCCAGGA |
| USP1 - F | CCAATGAGAGCGGAAGGAGG |
| USP1 - R | CACCAATTATATCTAGACCAAAGCC |

**Table 5. Primer sequences for qPCR to assess expression levels of selected genes**

From the differentially methylated genes related to CRC, four hypomethylated genes were selected to be validated for expression levels. The genes selected were compared in the dysplastic tissue and adjacent control sample to correlate the expression levels with their methylation.

## 3.12 In-vitro silencing of *SMAD7*

An attempt was made to knock down *SMAD7* mRNA in two colorectal cell lines and assess its effect on the transcription of previously identified genes (*CDH1*, *CGN* and *CLDN4*), which are related to epithelial-mesenchymal transition, and that of *DNMT1*.

### 3.12.1 Cell culture

DLD1 and HCT 116 colorectal cancer cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Sigma Life Science, Gillingham, Dorset, UK) supplemented with 10% (v/v) Fetal Calf Serum (FCS; Sigma) and 100 units/ml penicillin, 100 µg/ml

streptomycin, 2 mM L-glutamine (Sigma). Cells were cultured in T125 flasks or 6 well plates unless otherwise specified and maintained at 37°C in humidified 5% $CO_2$ incubator. Cells were routinely passaged when 70%-80% confluency was reached. To passage cells, the culture medium was aspirated, cells washed with PBS and trypsinised with EDTA-trypsin at 37°C for 3-5 minutes. Culture medium was added to a 1:1 ratio to inactivate trypsin and the cells were centrifuged at 1300 rpm for 3 minutes. The supernatant was discarded, and the cell pellet was suspended in the culture medium and split to the desired dilution.

### 3.12.2 Transfection of the cells with siRNA

DLD1 and HCT116 cell lines were cultured in 6 well culture plates. A seeding concentration was 100,000 cells/well in 2.5 mL of DMEM medium. After incubation in 37°C for 24 hours, the cells were transfected with *SMAD7* antisense (AS) and sense (S) oligonucleotide from ON-TARGET plus Human SMAD7-4092 siRNA-SMARTpool (Dharmacon #SO-2745395G) (Table6). Lipofectamine 2000 in Optimem medium with 1.2 ug/mL concentration of the AS and S oligonucleotide separately was used. Cells were harvested at different time points (24, 48 and 72 hours). Cells were processed for both RNA and protein extraction.

| Si-RNA | sequence |
|---|---|
| J-020068-05: Sense | GCUUUCAGAUUCCCAACUU |
| J-020068-06: Antisense | UGAAGGCGCUCACGCACUC |
| J-020068-07: Sense | CAUCAAGGCUUUCGACUAC |
| J-020068-08: Antisense | CCAGAUACCCGAUGGAUUU |

**Table 6. SMAD7 siRNA sequences**

Base sequences of the anti SMAD7 siRNA used to transfect the colon cancer cell lines with lipofectamine.

### 3.12.3 Transfection of cells with a pLKO-Tet-ON vector containing a shRNA

The pLKO-Tet-ON vector is a lentivirus-based vector generated by sequential PCR-based modifications. It contains all the required *cis*-elements for packaging, reverse transcription, and integration, which are required for subsequent production of the lentiviral particles. This vector also contains all the necessary components for the inducible expression of shRNA in the target cells. In the absence of tetracycline/ doxycycline, shRNA expression is repressed by constitutively expressed TetR protein. Upon the addition of tetracycline/ doxycycline to the growth media, shRNA expression is triggered resulting in target gene knockdown.

The shRNA oligos were custom designed using commercial software matching the non-alternating regions of the SMAD7 exon (Table 7).

| shRNA oligos | sequence |
|---|---|
| **Pair 1- Top** | ccggGAGGCTGTGTTGCTGTGAActcgagTTCACAGCAACAC AGCCTCtttttt |
| **Pair 1 - Bottom** | AATTaaaaaaGAGGCTGTGTTGCTGTGAActcgagTTCACAGC AACACAGCCTC |
| **Pair 2 - Top** | ccggGCTGCGGGGAGAAGGGGCGACctcgagGTCGCCCCTT CTCCCCGCAGCtttttt |
| **Pair 2 - Bottom** | AATTaaaaaaGCTGCGGGGAGAAGGGGCGACctcgagGTCGC CCCTTCTCCCCGCAGC |

**Table 7. SMAD7 ShRNA oligo sequences.**

Base sequences of shRNA pairs used to transfect the cell lines using the pLKO-Tet-ON construct.

*Oligo annealing*

The oligos were reconstituted in dd$H_2O$ to a concentration of 0.1 nmol/µl and mixed for a few seconds to dissolve. 11.25 ul of each oligo (Top and Bottom, 0.1 nmol/ul) and 2.5 ul 10X annealing buffer (1M NaCl, 100 mM Tris-HCl, pH=7.4) were mixed in

a separate 1.5 ml tube. 200 ml of water was brought to boiling temperature and the beaker was placed on an 1.5 ml tube rack on the bench (having the rack between the beaker and the bench slows down the cooling process). The temperature of the water was maintained at 95$^{o}$ C. The oligo mixtures were placed in a water bath and allowed to cool down naturally to 30$^{o}$C (2-3 hours). One ul of the oligo mixture was diluted 1:400 in 0.5X annealing buffer (1 µl + 399 µl of 0.5X buffer) and stored at -20$^{o}$C.

*Ligation of the oligos to pLKO-Tet-ON*

The vector pLKO-Tet-On was digested with AgeI (New England Biolabs (NEB) #R0552S) and EcoRI (NEB #R0101S). Six µg of pLKO.1 TRC-cloning vector were mixed with 5 µl 10x NEB buffer 1 µl, AgeI 1 µl, ddH2O up to 50 µl, incubated at 37°C for 2 hours and purified using Qiaquick gel extraction kit (Qiagen) to a 30 µl. The elute was digested with EcoRI mix containing 5 µL 10x NEB buffer for EcoRI, 1 µl EcoRI to 14 µl ddH2O and incubate at 37°C for 2 hours. The digested DNA was run on 0.8% low melting point agarose gel until 2 bands were seen distinctly, one at 7kb and one 1.9kb (Figure 11). The cut out 7kb band was placed in a sterile microcentrifuge tube and the digested vector was gel purified (the stuffer is released as a 1.8 kb fragment and was discarded).

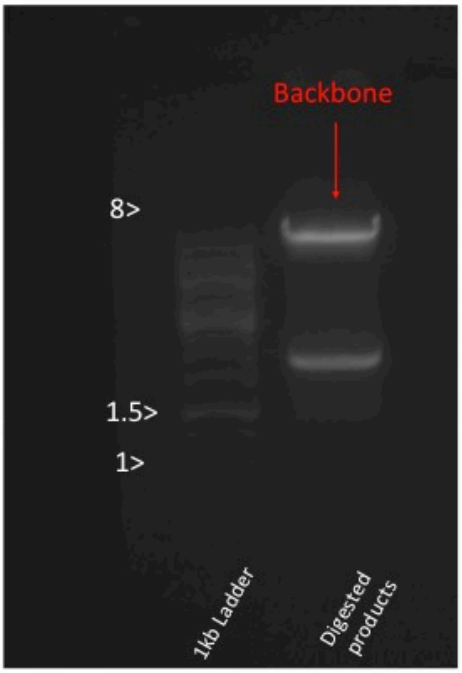



**Figure 11. Separation of pLKO backbone.**

Gel electrophoresis of the digested pKLO showing the backbone and the stuffer separately. The backbone was cut out from the gel and purified for the ligation of shRNA oligos. The T4 ligation reaction was set up as shown in table 8.

| Component | volume |
|---|---|
| **Oligo dilution (or 0.5X buffer as a negative control)** | 1 µl |
| **Gel-purified digested pLKO-Tet-On (10-20 ng)** | 1 µl |
| **10X ligase buffer** | 1 µl |
| **T4 DNA ligase** | 1 µl |
| **Water** | 6 µl |
| **Total** | 10 µl |

**Table 8. Components mix for the T4 ligation reaction.**

*Transformation of the ligated product*

Transformed competent Stbl3 *E. coli* cells were mixed with 4 ul of the total ligation reaction volume, following established procedures (pre-chill ligation reaction on ice, add 4 µl of ligation reaction to E.coli, incubation for 20-30 min, heat shock for 30-45 sec at 42°C, incubate on ice for 5 min, add 500 ul of SOC growth media, incubate at 37°C while shaking for 30 min). 200 µl of *E.coli* were plated on to pre-warmed LB plates containing 50µg/mL of ampicillin (LB/Amp) and incubated for 24 hours. The distinct colonies were manually selected and miniprep cultures (5-6 ml of LB/Amp) were inoculated. The DNA was isolated using Qiagen miniprep kit (Qiagen #27104) and the elute was subjected to XhoI (Thermo Scientific) restriction digest reaction as shown in table 9.

| Component | volume |
|---|---|
| **miniprep DNA** | 1 µl (4 µg) |
| **10x buffer R** | 2 µl |
| **Xho1** | 2 µl |
| **$H_2O$** | 15 µl |
| **Total** | 20 µl |

**Table 9. Component mix for restricted digestion.**

The products were gel electrophoresed in a 1.5% agarose gel. The colonies with two migrating bands near 200bp were selected as successfully ligated colonies (Figure 12). Selected colonies were cultured in 100ml of LB broth medium overnight. The plasmid was purified using the QIAprep miniprep kit and quantified with Qubit fluorometric quantification.

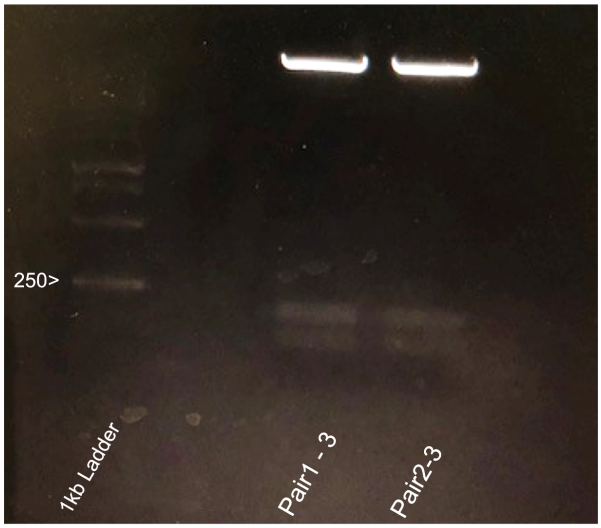



**Figure 12. Selection of transformed E. *coli* colonies.**

Gel electrophoresis of digested products in 1.5% agarose gel from the selected bacterial colonies indicating two distinct bands at and below 200 bp region indicating ligated oligos.

### 3.12.4 Viral transfection

*Co-transfecting HEK293T*

$1.3x10^6$ HEK293T cells were seeded onto 6 cm dishes in complete DMEM, so that ~90% confluency was reached after 24 hours. On day 2, 5 µg of transfer vector, 1.75 µg of pMD2.G and 3.25 µg pCMVΔR8.74 were mixed in a 1.5 ml tube (max volume of DNA = 500ul per reaction). Then, 30ul of FugeneHD was incubated with 500 µl of Optimem for 10 min before allowing the incorporation of the purified DNA for at 45 min.

After this time, the DNA/Fugene/Optimem mix was added to HEK cells, supplied with fresh media, dropwise, and the cells were left to incubate overnight 16-24h max. After 16-24h, the media on HEK cells was changed.

*Viral harvesting*

After 48h, the viral supernatant was centrifuged for 5 min at 1500rpm in room temperature (inside the hood), filtered using a 0.45 µm filter and stored at -80C.

*Transduction*

HCT116 and DLD1 cells were plated in a 6-well plate (100,000 cells/well) aiming for 50% confluency on the following day. Polybrene was added at 1:1000 concentration to filtered viral supernatant and mixed. The media was aspirated from both treatment wells and controls and 500 µl of viral supernatant was added onto cells, while normal medium was added to controls for 'mock transduction'. The media was replaced after 24h hours and the supernatant was disposed of after disinfection. After 72h, selection of all cells (transduced and controls) was started by adding puromycin (Thermo Fischer # A1113803) at a concentration of 1µg/ml. The mock transduced plate was used to continually assess cell killing by light microscopy and the media was changed every 48 hours. Cells were split as necessary until all the cells in the mock transduced well were dead.

After the selection, doxycycline was used at 10 ng/ml concentration to induce the expression of shRNA on Tet-ON. Cells were harvested at 24 and 48-hour intervals to assess for the concentration of SMAD7mRNA levels.

### 3.12.5 RNA extraction

RNA extraction was performed using the miRNeasy Mini RNA purification kit (#217004). First, 700 µL of Trizol (Qiazol©, Qiagen, Hilden, Germany) was directly added to the wells. The lysate was mixed till it was homogenous and 140 µL of

chloroform was added under a fume hood. After mixing well the tubes were centrifuged at 12,000 x g at 4°C for 15 minutes. The upper aqueous phase was transferred into a new tube and 1.5 volumes of 100% ethanol was added. The sample was pipetted into a miRNeasy Mini spin column and centrifuged at 8000 g for 30s at room temperature. DNAse digestion was performed by adding 10 µL of DNAse I and 70 µL of RNase-Free DNase directly on to the membrane after a single wash with 350 µL of Buffer RWT (a stringent washing buffer). This step was followed by a single wash with Buffer RWT and two washes with Buffer RPE (washing membrane-bound RNA). RNA was extracted with 30 µL RNase-free water.

### 3.12.6 Primer annealing temperatures and standard curves for *SMAD7* primers

To determine the optimised annealing temperature of the primers, a gradient PCR reaction was performed on the cDNA samples derived from the DLD1 cell line, using Taq DNA Polymerase (Qiagen). PCR conditions were as follows: initial denaturation at 94$^{o}$C for 5min; amplification of 34 cycles at 94$^{o}$C for 30sec (denaturation), 54$^{o}$ C to 65$^{o}$ C for 40sec (annealing), 72$^{o}$ C for 30sec (elongation); final elongation at 72$^{o}$C for 10min. This was followed by electrophoresis on a 1.5% agarose gel with a DNA Hyperladder 100bp (Bioline, London, UK) to verify the presence of a single amplified template and the desired annealing temperature, which was then set at 63$^{o}$C (Figure 13).

To determine primers efficiency and the necessary concentrations of primers to be loaded in the future reactions, qPCR reactions at the optimised annealing temperature were performed with serially diluted concentration of the individual primers (10, 5, 1, 0.5μM) and the cDNA samples 20 ng/μL, and both the standard curve and the melting curve was analysed on the StepOne Plus Real-Time PCR System software (Applied Biosystems, MA, USA) (Figure 14). Samples were loaded in three technical triplicates in all cases. qPCR conditions were as follows: 20sec at 95$^{o}$C; 3sec at 95$^{o}$C and 30sec at 63$^{o}$C (40 cycles); 15sec at 95$^{o}$C. Primer efficiency slopes were compared to assure there was no significant difference amongst them and melting curve was analysed to determine the absence of multiple peaks, which could be formed by primer non-specificity or primer dimer formation.

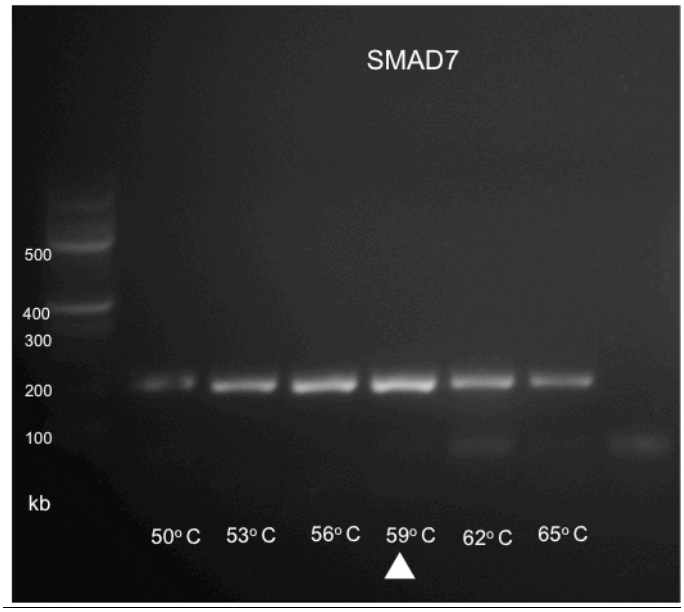



**Figure 13. Optimum annealing temperature of SMAD7 primer.**

Gel electrophoresis of PCR products of SMAD7 primers at different annealing temperatures. The most prominent single band was seen at 59$^{o}$ C which was chosen as the annealing temperature for subsequent qPCR cycles.

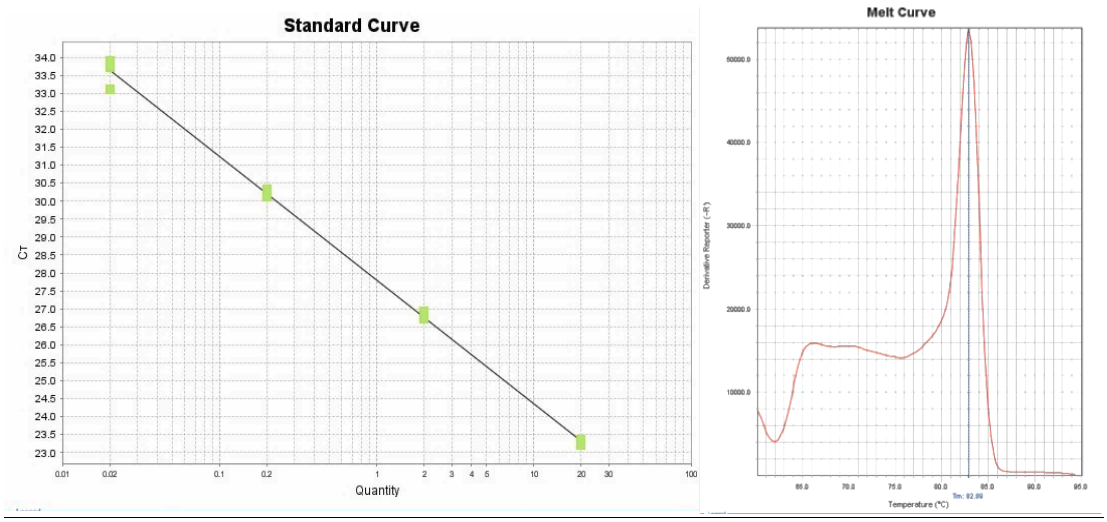



**Figure 14. Validation of primer efficacy.**

The standard curve produced with 10 fold dilutions of template DNA producing a curve with a $R_0$ of 0.99 indicating the efficacy of the primers. The melt curve with a single distinct peak indicating the PCR products are of a single type.

### 3.12.7 Reverse transcription and real time PCR

The cDNA was obtained from extracted RNA using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific) according to manufacturer's instructions. Briefly, RNA samples were mixed with 4µl of 5X Reaction Mix (Reaction Buffer, dNTPs, oligo (dT)18 and random hexamer primers), 2µl of Maxima Enzyme Mix (Reverse Transcriptase, RNase Inhibitor), and nuclease-free water to obtain a total volume of 20µl in each PCR tube. The conversion reaction was performed in a thermocycler (DNA Engine Dyad, Bio-Rad) at 25$^{o}$C for 10min, followed by a heating step at 85$^{o}$C for 5min to terminate the reaction. After cDNA conversion, a 1.5% agarose gel was run with Reverse transcriptase minus (RT-) negative control and no template control (NTC) to ensure no contamination was present.

### 3.12.8 Protein extraction

The cells were lysed via incubation with 70 µl of Radioimmunoprecipitation assay (RIPA) buffer on ice for 15 minutes. The lysate was then centrifuged at 15000 RPM in 4°C for 15 minutes.

### 3.12.9 protein quantification

Protein quantification was done using the Bio-Rad DC (Bio-Rad,CA, USA) microplate assay technique. Serial dilutions of the protein standard were made ranging from 0.2mg/ml to 1.6mg/ml using bovine albumin. Reagent mix was prepared by mixing reagent A with reagent S at a 1:50 ration. 5 µl of the standard protein dilutions and the extracted proteins were added to the micro-plate in triplicates. Each well was then added with 25 µl of the reagent A and 200 µl of reagent B. The plate was placed on a plate mixer for 5 seconds. Any bubbles that were visible were popped with a clean pipette tip. After 15 minitues the plate was kept in a Sunrise® ( TECAN, Swizerland) micro-plate reader with flurometry settings at 750 nm. The average absorbance values were used to create the standard curve and the concentration of the extractions was measured in relation to the curve.

### 3.12.10 Western blot

Extracted proteins were quantified using the Bradford assay with Tecan Spectrophotometer. Samples were prepared with 20µg of proteins per well at a 40µl total volume. Each sample included 10µl of sample buffer (Novex, Bolt #B0007, Thermo Fisher Scientific), 4 µl of oxidizing agent (Novex, Bolt #1772697) and $D_2O$ up to a total of 40 µl. The samples were heated up to $70^0$C for 10 mins.

Novogen Tris-Bis 10% (Thermo Fisher Scientific) were used with Mini Gel Tank (Life Technologies). NuPAGE™ MOPS SDS Running Buffer (20X) diluted to 1X was used at a running voltage of 130 mV for 40 minutes.

Transfer buffer was freshly prepared with Tris 3.03g, glycin 14.4g, methanol 200 ml in 800 ml of distilled water. The gel was transferred onto the nitrocellulose membrane, packed tightly with sponges on either side and the transfer chamber was loaded with 500 ml of transfer buffer. The chamber was left on ice and the running voltage was 100 mV for 1 hour.

The membranes were washed in TBST (Tris Buffered Saline, with Tween 20) for 5 mins and was blocked in 5% milk (5 g of milk powder in 0.05% PBST) for 2 hours. The primary antibody for SMAD7 was applied in 1:100 dilution in milk and left shaking overnight at $4^0$C. After three washes with 0.05% PBST, the HRP linked anti-Mouse – 1:1000 (CST – 7076S) secondary antibody was applied for 1 hour before developing in the dark room.

# 4.0 Results: 1

## 4.1 Methylation landscape of colitis associated neoplasia

### 4.1.1 Scope of the chapter

In the background of consistent genetic markers to predict the development of CAC, focus has shifted to epigenetics. However, the methylation landscape of the colon, from the inflamed stage to the development of colitis-associated cancer has not been well studied up to now. One of the aims of this study was to test the feasibility of applying whole genome bisulphite sequencing to identify a predictive marker for CAC, present from the early non-neoplastic stage of the disease.

Therefore, to identify potential early biomarkers to predict CAC, the methylation pattern from different stages of UC associated colon cancer was examined using whole genome bisulphite sequencing of laser captured epithelial cells. Each sample was matched with an adjacent tissue or blood control and the methylation pattern in normal colonic epithelium was subtracted from the data from the diseased tissue. From the identified common methylation markers, those linked to CRC were identified.

### 4.1.2 Chapter summary

The chapter describes the general methylation pattern of the sequenced samples and specifically the identified common markers for both promoter and gene body methylation patterns. Sequencing data has been validated for a subset of these genes with expression levels with RT-qPCR. This chapter describes potential genes and their methylation patterns that can be used as potential biomarkers to be validated in a prospective study.

### 4.1.3 Sample characteristics

Eight samples were subjected to whole genome bisulphite sequencing (WGBS) next generation sequencing (NGS) using the Illumina 4000 platform. The test samples consisted of DNA isolated from laser-captured microdissected epithelial cells from non-inflamed (healthy), inflamed, dysplastic and colitis associated cancer tissue. The matched controls were DNA extracted from buffy coat for the non-inflamed and inflamed tissue and DNA extracted from adjacent colonic mucosal cells for the dysplastic and cancer tissue. The details of the patients from which the samples were obtained are given below (**Table 6**)

| Sample name | Type | Fixation | Age | Sex | Disease duration | Site |
|---|---|---|---|---|---|---|
| ST001-1 | Non-inflamed | Fresh frozen | 22 years | Female | N/A | Sigmoid |
| ST001-2 | Buffy coat | | | | | |
| ST002-1 | Inflamed | Fresh frozen | 29 years | Male | 10 years | Rectum |
| ST002-2 | Buffy coat | | | | | |
| ST004-A | Dysplastic | Fresh frozen | 71 years | Female | 20 years | Rectum |
| ST004-C | Non dysplastic-adjacent | | | | | |
| 43795-5A | Colitis associated cancer | FFPE | 64 years | Female | 34 years | Colon |
| 43795-A12 | Adjacent to cancer | | | | | |

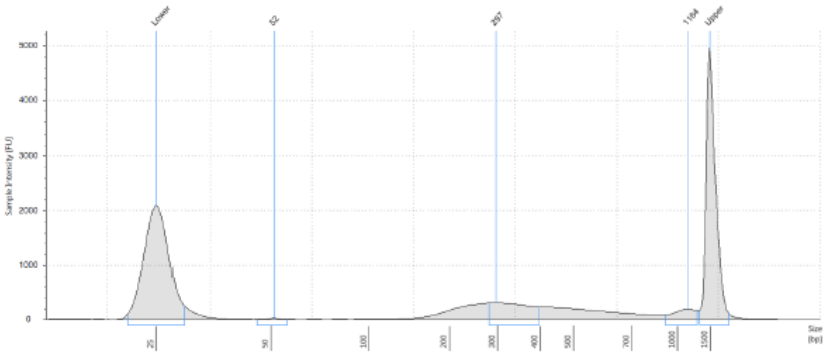
**Table 10. Details of the samples selected for NGS**

Colonic epithelium from patients with the inflammatory phase of UC, dysplasia due to long standing UC and CAC were selected. The adjacent non neoplastic mucosa was selected as the control sample for the dysplasia and cancer tissue. For the inflammatory mucosa, the buffy coat was used as a control sample to eliminate the age-related methylation changes. The normal mucosa was from a patient who did not have an inflammatory condition was also matched with his buffy coat.

### 4.1.4 Quality of library preparations

Libraries were prepared for sequencing using the PicoMethyl kit (Zymo) using starting amount of 100 ng of DNA. Quality of the libraries is assessed based on the fragment lengths and the distribution of the lengths in the total sample. The library quality assessed using Tapestation 2000 matched with the validation curve provided by the manufacturer with a majority of fragment sizes range from 150 bp – 500 bp. (**Figure 11, a. and b.**).

a)

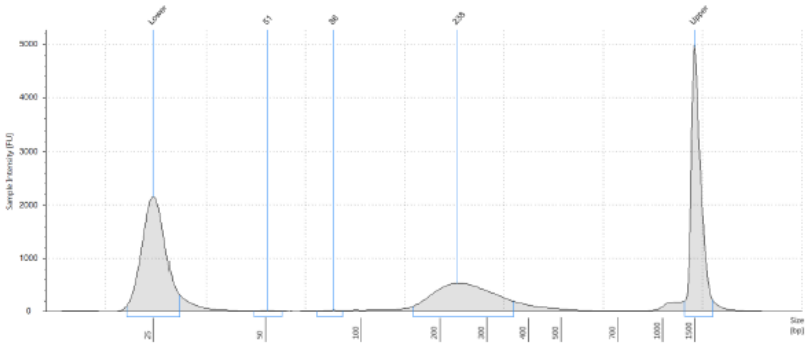



**Sample Table**

| Well | Conc. [ng/µl] | Sample Description | Alert | Observations |
|---|---|---|---|---|
| B1 | 2.30 | 004A | | |

**Peak Table**

| Size [bp] | Calibrated Conc. [ng/µl] | Assigned Conc. [ng/µl] | Peak Molarity [nmol/l] | %Integrated Area | Peak Comment | Observations |
|---|---|---|---|---|---|---|
| 25 | 7.23 | - | 445 | - | | Lower Marker |
| 52 | 0.0405 | - | 1.20 | 1.76 | | |
| 297 | 1.69 | - | 8.75 | 73.17 | | |
| 1164 | 0.578 | - | 0.764 | 25.07 | | |
| 1500 | 6.50 | 6.50 | 6.67 | - | | Upper Marker |




**Sample Table**

| Well | Conc. [ng/µl] | Sample Description | Alert | Observations |
|---|---|---|---|---|
| C1 | 4.44 | 0736-A17 | | |

**Peak Table**

| Size [bp] | Calibrated Conc. [ng/µl] | Assigned Conc. [ng/µl] | Peak Molarity [nmol/l] | % Integrated Area | Peak Comment | Observations |
|---|---|---|---|---|---|---|
| 25 | 6.96 | - | 428 | - | | Lower Marker |
| 51 | 0.0351 | - | 1.05 | 0.79 | | |
| 86 | 0.0461 | - | 0.828 | 1.04 | | |
| 238 | 4.36 | - | 28.3 | 98.17 | | |
| 1500 | 6.50 | 6.50 | 6.67 | - | | Upper Marker |

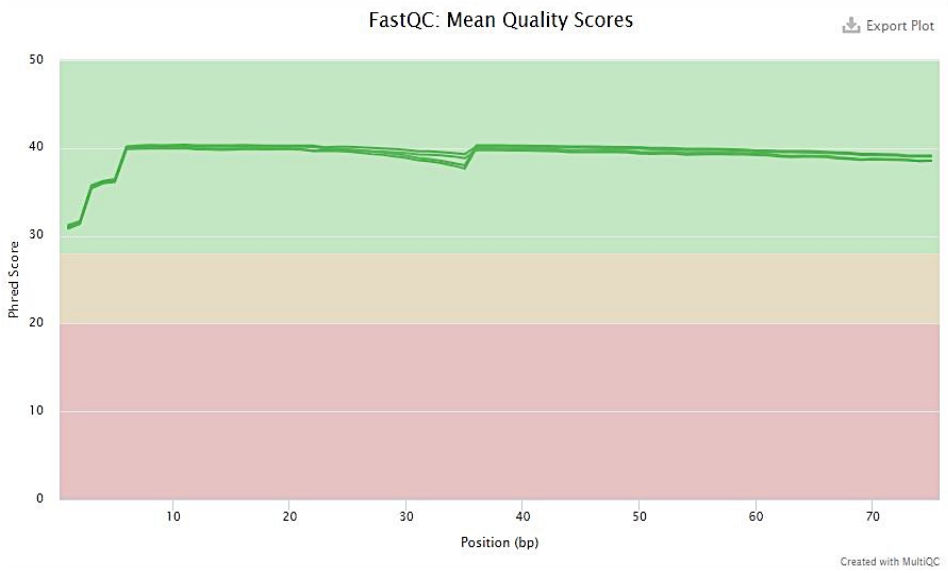
**Figure 15. Quality of the library preparations**

Library preparation quality assessment for tissue sample with dysplasia (a) and adjacent tissue (b), on Agilent Tapestation 2000. The shape of the curve matched the recommendation by the manufacturer. Majority of the library fragments are less than 300 bp in length.

### 4.1.5 Quality assessment of sequencing data

All samples passed the Fast QC quality assessment in both the mean quality scores and per base sequence quality score. All predicted scores were in the acceptable range (**Figure 12**).

a




b

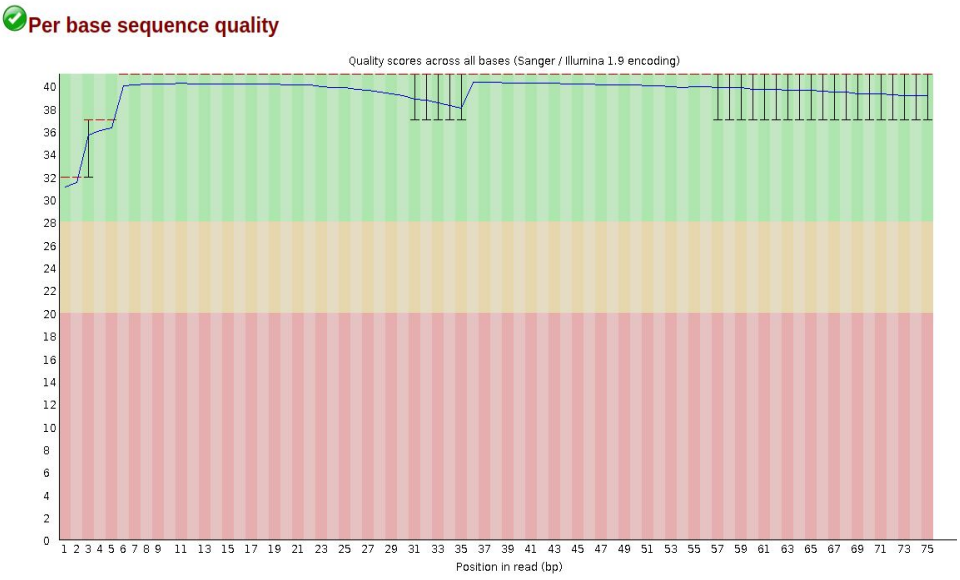



**Figure 16. Quality score for each base position in 75 base read lengths**

a) Mean predicted quality score for 75 base read lengths showing all reads within the desired 'green' zone. b) Values for individual base positions indicating quality scores between 32 and 40. Values are representative for 4 samples.

## 4.3 Analysis of promoter methylation in different stages of CAC

### 4.3.1 General distribution of promoter methylation

Methylation in mammalian DNA occurs predominantly (not exclusively) at cytosine bases. A CpG island is a region where a large number of repeated cytosine-phospho-guanine (CpG) dinucleotide sequences are clustered together. Around 60 - 80% of the CpG residues are methylated in mammalian DNA, and the non-methylated CpG elements tend to cluster together as islands. These islands are generally around 200-300 bp in length and are found adjacent to promoter regions or other regions implied in transcription regulation. DMRs are genomic regions with different DNA methylation patterns, across different biological samples. These usually have a functional importance such as transcription regulation.

Each sample was compared to its control – either adjacent colonic epithelium or buffy coat from the same patient to exclude the age-related methylation changes. The number of DMRs in promoters for each sample-control pair is as follows (**Table 7**).

| Sample | Control | Number of DMR |
|---|---|---|
| **ST001-1: Non inflamed** | ST001-2, Buffy coat | 5770 |
| **ST002-1: UC, Inflamed** | ST002-2, Buffy coat | 5505 |
| **ST004-A: UC, Dysplasia** | ST004-C: Adjacent | 2190 |
| **43795-5A UC, Cancer** | 43795-12A: Adjacent | 6613 |

**Table 11. Total number of differentially methylated regions in test samples.** Each test sample was matched with a control form the same patient. For the neoplastic tissue the adjacent non-neoplastic mucosa was used while for the inflamed and the normal control their buffy coat was used.

The distribution of the DMRs in different tissue types (non-inflamed/ non diseased, inflamed, dysplastic and cancer) was analysed (**Figure 13**). The M value was plotted against the gene identification to observe the spread of the DMRs across the genomes. The M value was chosen to compare methylation levels as it has been proposed to be more compatible with the statistical analysis as its range is not bound to a certain value. In the traditional Beta-value the range is limited to between 1 and 0, and the violates the Gaussian-distribution assumption made by most statistical tests. The two values have been compared in their ability to more accurately predict the true level of methylation (197). The relationship between Beta-value and M-value is a sigmoid curve (197). The curve flattens out towards values closer to 0 and 1 in Beta-value, which is around 6 in M value. Therefore during conversion methylation levels closer to 0 or 1 in Beta-value can be polarised to extreme ends in the M-value as seen in the plot.

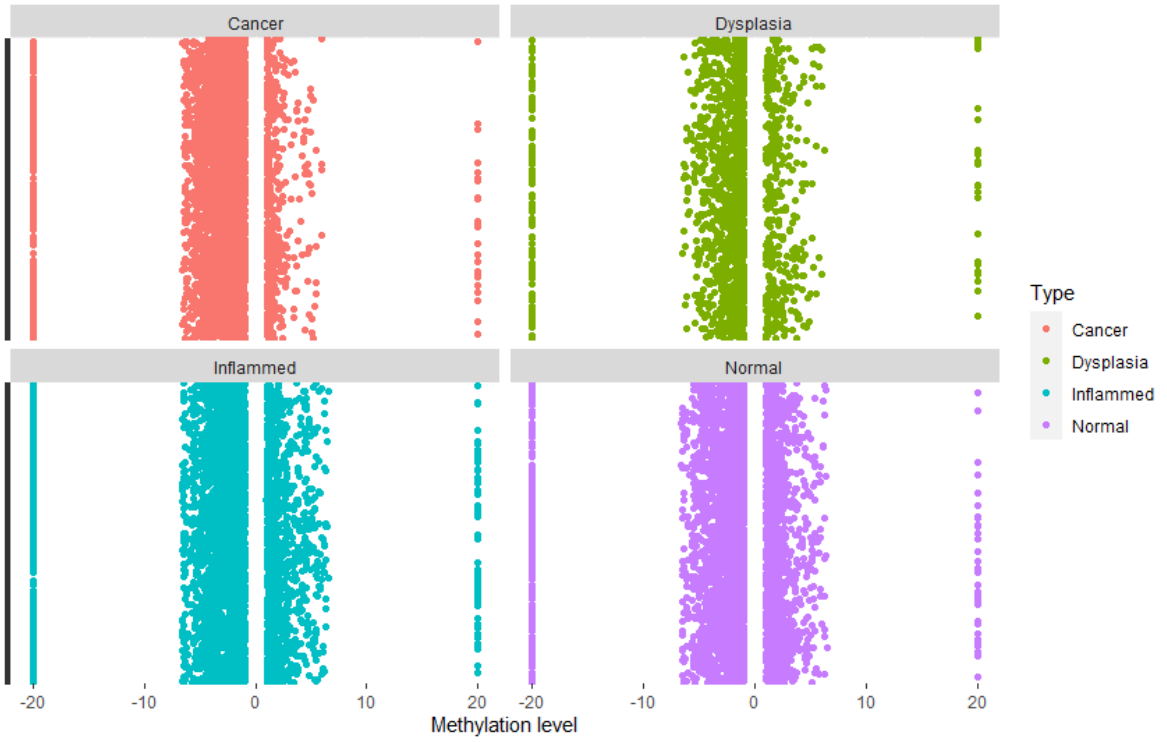



**Figure 17. DMRs at promoter regions in different stages**

Scatter plot comparing the general distribution of the methylation pattern in different stages of CAC. Compared to the normal and inflamed samples more hypomethylated promoters are observed in the dysplastic and cancer tissue.

Dysplasia and cancer show a higher proportion of hypomethylated promoters (compared to normal and inflamed tissue. Overall 91% (total n= 6080) of the DMRs in cancer tissue and 80% (total n=2007) of the DMRs in the dysplastic tissue were hypomethylated compared to the control samples. Normal tissue sample demonstrated 76% (total n=5359) of the DMRs to be hypomethylated while the inflamed tissue had 78% (total n=5093). This is keeping with the general pattern identified in cancers. A significant increase in the hypomethylated proportion was noted from dysplasia to cancer ($p<0.001$) and from inflammation to dysplasia ($p<0.001$) while the difference in hypomethylated promoters as a proportion of all

DMRs in inflamed tissue in comparison to the normal epithelium was non-significant (p=0.631).

### 4.3.2 Pathway annotation of genes with differentially methylated promoters

The genes showing differential methylation were annotated and their pathway enrichment was analysed. The top 10 ranked pathways where the differentially methylated genes are represented were identified. The representations were related to same top 3 pathways in all three tissue types. Highest representation in all three stages were related to metabolic pathways, pathways in cancer and PIK3-Akt signalling pathway (**Figure 14**).

### 4.3.4 Identifying promoter methylation pattern across all stages of CAC

As the first step to identify potential methylation markers that influence the process of CAC I looked at promoter methylations present across all stages from inflammation to cancer and not present in normal colonic epithelium. For any molecular marker to be used as a predictor it needs to be consistently present from the early stages of the disease. The workflow is outlined in **figure 15**.

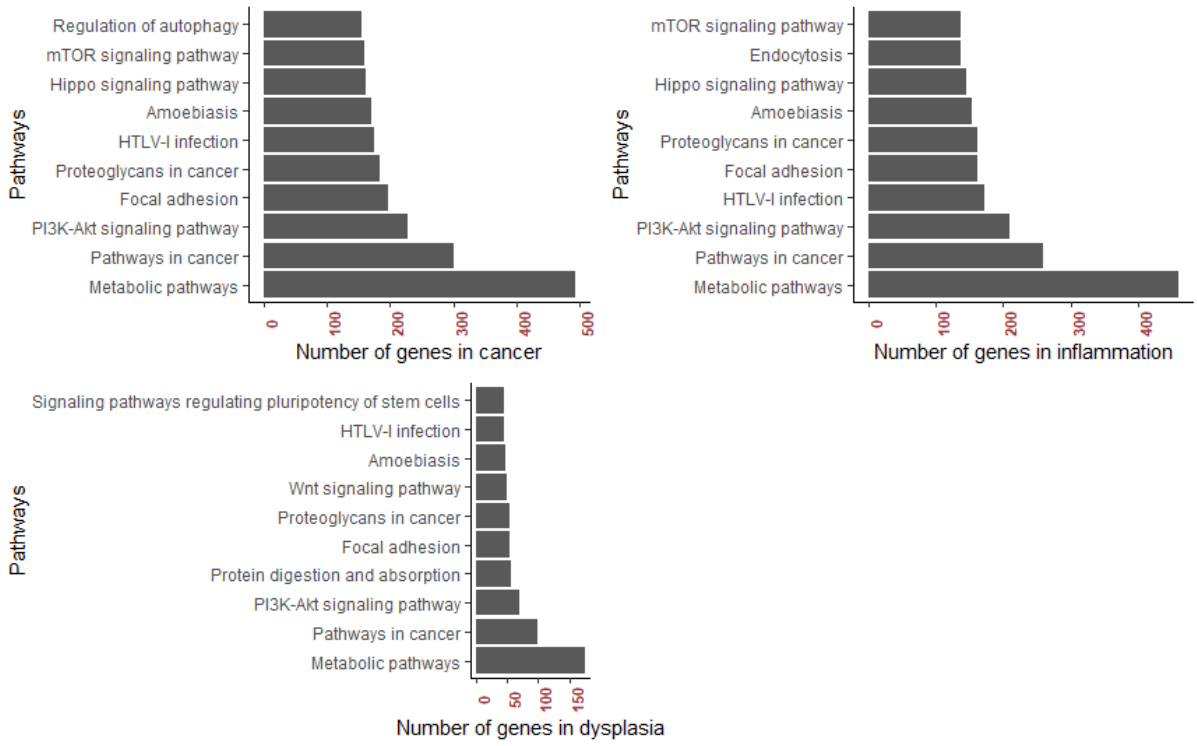



**Figure 18. Annotation of DMRs to pathways in different tissues**

The genes with differentially methylated promoter regions (DMR) were annotated to pathways. Annotated pathways are arranged in the descending order according to the number of genes annotated to them. The highest numebr of genes were annotated to metbolic pathways, pathways in cancer and PI3K-Akt signalling pathway in all three stages.

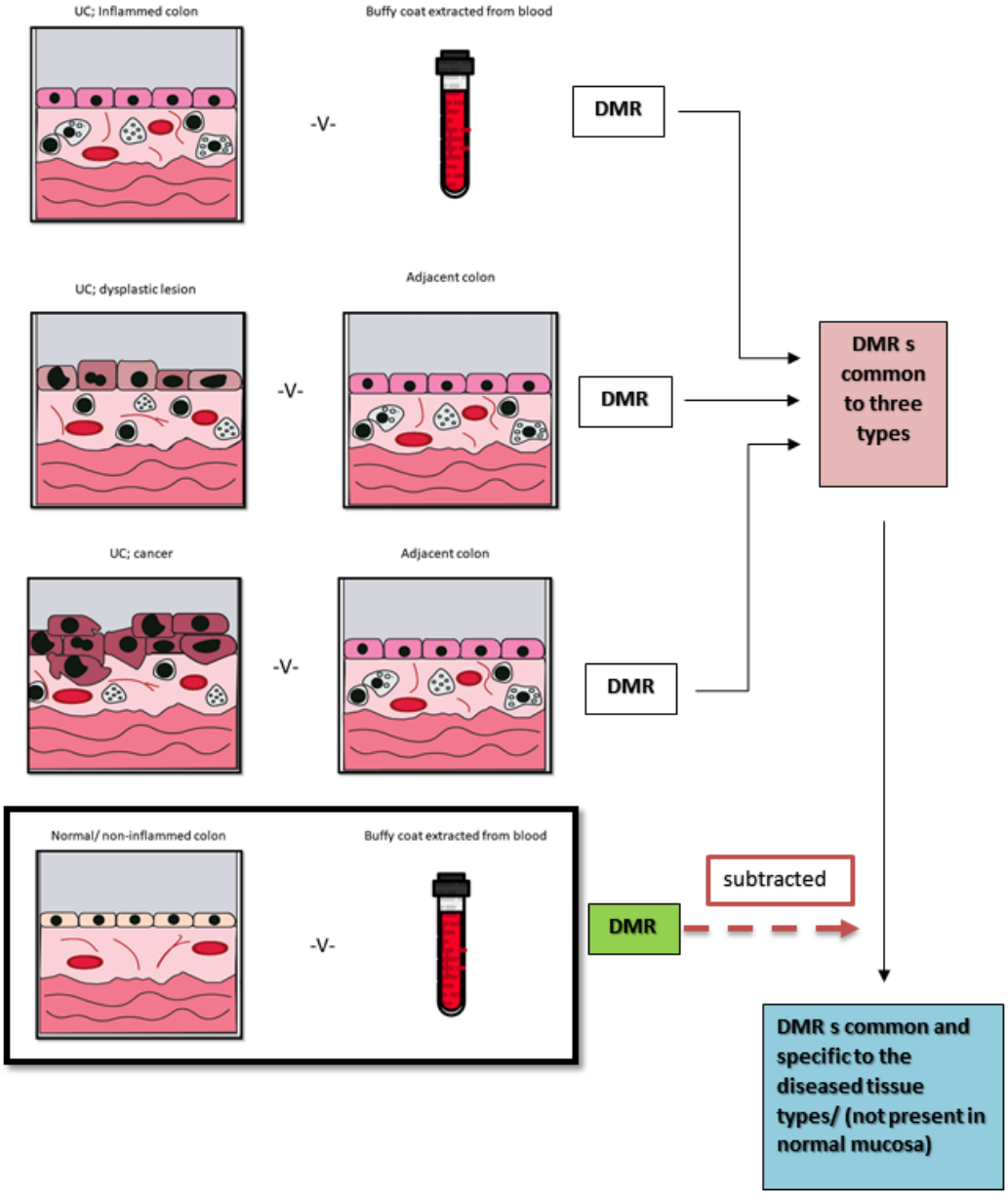



**Figure 19. Analysis pathway of DMRs between different tissue types**

Analysis pathway to detect common methylation markers across tissue types and not present in normal colonic mucosa. The methylation pattern in each stage is matched to a control (either adjacent tissue or blood) to subtract the age related methylation. The common DMRs consistently present over inflamed, dysplastic and cancer stages were recognised. All DMRs detected in normal colonic mucosa were then subtracted

from the common list leaving only those consistently present in the UC affected inflamed, dysplastic and cancer epithelia.

The hyper and hypomethylated DMRs in each sample were separated based on the M value (log2ratio) using the following codes in base R.

To separately identify the hyper and hypomethylated promoters in inflamed tissue, new data frames (DMR_INF_pos – hypermethylated and DMR_INFneg – hypomethylated) were formed based on the M value (methylation level);

```
DMR_INF_pos<-DMR_INF[DMR_INF$log2Ratio>0,]
DMR_INF_neg<-DMR_INF[DMR_INF$log2Ratio<0,]
```

Similarly, the DMRs in dysplastic (DMR_DYS) cancer (DMR_CA) and normal (DMR_NORM) tissues were also separated based on the M value (log2ratio).

```
DMR_DYS_pos<-DMR_DYS[DMR_DYS$log2Ratio>0,]
DMR_DYS_neg <- DMR_DYS[ DMR_DYS$log2Ratio<0,]
```

```
DMR_CA_pos<- DMR_CA_[ DMR_CA $log2Ratio>0,]
DMR_CA_neg  <-DMR_CA[ DMR_CA$log2Ratio<0,]
```

```
DMR_NORM_pos<- DMR_NORM [ DMR_NORM $log2Ratio>0,]
DMR_NORM_neg<- DMR_NORM [ DMR_NORM $log2Ratio<0,]
```

Following the separation, the hyper and hypomethylated regions common to all there stages (inflammation/ dysplasia/ cancer) were selected using the ***cbind*** function in R base package:

```
Combined_DMR_INF_DYS_pos<cbind(DMR_INF_pos[DMR_INF_pos$GeneID%in%DMR
_DYS_pos$GeneID,],DMR_DYS_pos[DMR_DYS_pos$GeneID%in%DMR_INF_pos$Gene
ID])
```

```
combined_DMR_pos <-
cbind(DMR_CA_pos[DMR_CA_pos$GeneID%in%Combined_DMR_INF_DYS_pos,],Com
```

```
bined_DMR_INF_DYS_pos[Combined_DMR_INF_DYS_pos$GeneID%in%DMR_CA_pos]
)
Combined_DMR_INF_DYS_neg<cbind(DMR_INF_neg[DMR_INF_neg$GeneID%in%DMR
_DYS_neg$GeneID,],DMR_DYS_neg[DMR_DYS_neg$GeneID%in%DMR_INF_neg$Gene
ID])
```

```
combined_DMR_neg<-
cbind(DMR_CA_neg[DMR_CA_neg$GeneID%in%Combined_DMR_INF_DYS_neg,],Com
bined_DMR_INF_DYS_neg[Combined_DMR_INF_DYS_neg$GeneID%in%DMR_CA_neg]
)
```

From the two data sets of common hyper and hypomethylated promoter regions the ones observed in the normal colon sample were deducted. The function %notin% to negate a selection was defined and used in *dyplr* package.

```
Specific_DMR_pos<-combined_DMR_pos[combined_DMR_pos$GeneID %notin%
promoter_NORM_pos$GeneID,]
```

```
Specific_DMR_neg<-combined_DMR_neg[combined_DMR_neg$GeneID %notin%
promoter_NORM_neg$GeneID,]
```

This revealed 6 hypermethylated promoters common to all stages and 63 hypomethylated promoters common to all three stages (**Tables 8 & 9**) that were not observed in the normal colonic epithelium.

Six genes were found to have methylated promoters in all three stages of CAC and not detected in normal colonic epithelium. These genes were exclusive to the disease-affected epithelium (**Figure 16,17 & 18**).

| Gene ID | Function | Inflammation M value | Dysplasia M value | Cancer M value |
|---|---|---|---|---|
| IGF2 | Protein coding | 1.9 | 1.8 | 1.7 |
| ICAM5 | Protein coding | 1.5 | 20.0 | 2.6 |
| WT1 | Protein coding | 2.9 | 4.3 | 2.1 |
| SLC6A5 | Protein coding | 2.1 | 1.2 | 1.1 |
| ARMT1 | Protein coding | 1.0 | 20.0 | 5.4 |
| MIR124-2 | NC-RNA | 3.6 | 1.1 | 1.0 |

**Table 12. Hypermethylated promoters common across all stages of CAC**

These six genes were all hypermethylated in the inflamed, dysplastic and cancer epithelium compared to their coupled controls but not in the normal colonic epithelium.

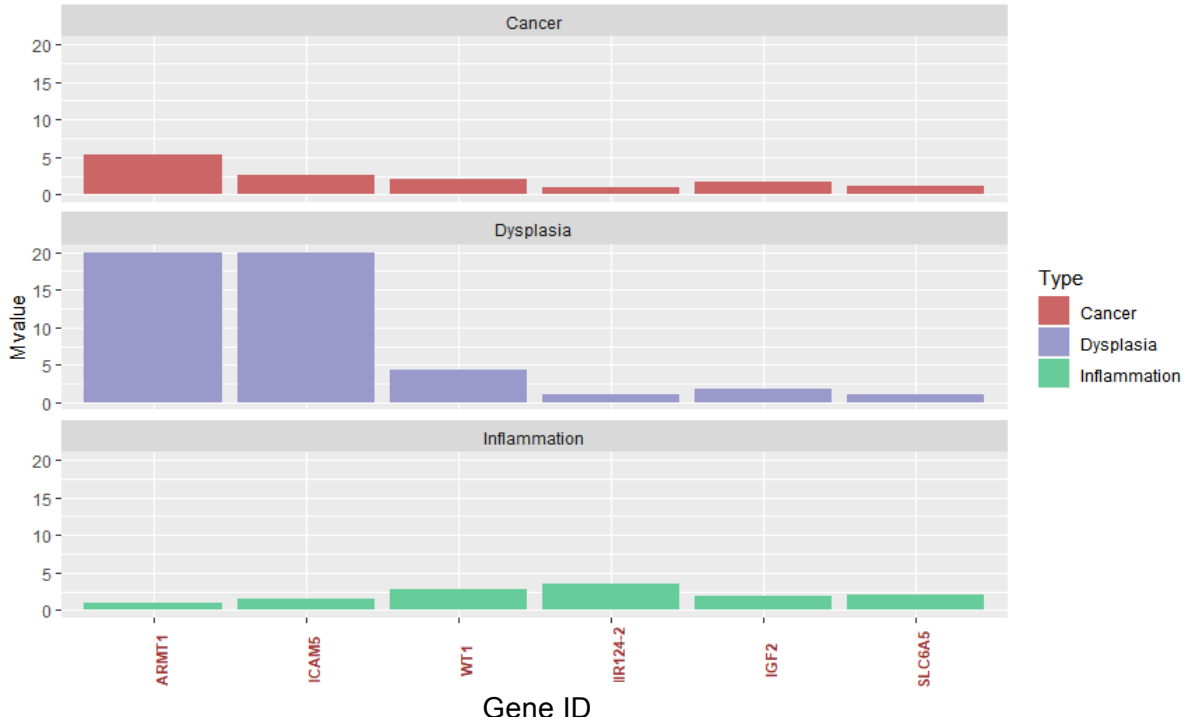



**Figure 20. Hypermethylated promoters in different tissue types**

Hypermethylated promoters are compared to control samples (blood/ adjacent tissue) and arranged in order. From the 6 genes identified 4 were linked to CRC previously.

| Gene ID | Function | Inflammation M value | Dysplasia M value | Cancer M value |
|---|---|---|---|---|
| CXADR | Protein coding | -1.4 | -1.2 | -2.6 |
| DHCR24 | Protein coding | -2.8 | -1.8 | -1.7 |
| ELAVL4 | Protein coding | -5.0 | -1.5 | -1.1 |
| FCER1G | Protein coding | -2.8 | -1.1 | -3.5 |
| HK1 | Protein coding | -4.6 | -20.0 | -1.1 |
| EIF6 | Protein coding | -1.1 | -1.1 | -1.0 |
| LFNG | Protein coding | -1.6 | -2.4 | -1.7 |
| TACSTD2 | Protein coding | -2.8 | -1.3 | -5.6 |
| MAGOH | Protein coding | -1.4 | -1.6 | -2.5 |
| MAP4 | Protein coding | -1.3 | -2.1 | -1.3 |
| MAP3K5 | Protein coding | -3.4 | -5.0 | -1.4 |
| NTN3 | Protein coding | -1.5 | -2.1 | -1.5 |
| PRDX1 | Protein coding | -20.0 | -1.2 | -1.1 |
| PCK1 | Protein coding | -1.3 | -2.4 | -2.2 |
| SRGN | Protein coding | -4.9 | -3.2 | -1.4 |
| PSMB2 | Protein coding | -5.5 | -1.0 | -1.1 |
| RLF | Protein coding | -2.9 | -1.4 | -1.1 |
| S100A12 | Protein coding | -20.0 | -1.7 | -2.4 |
| TCEA2 | Protein coding | -1.0 | -1.3 | -2.9 |
| USP1 | Protein coding | -4.5 | -2.6 | -2.1 |
| ALX1 | Protein coding | -20.0 | -5.0 | -3.7 |
| RBM10 | Protein coding | -1.8 | -1.1 | -2.9 |
| DIRAS3 | Protein coding | -5.3 | -1.5 | -3.7 |
| CER1 | Protein coding | -1.2 | -1.4 | -1.3 |
| DNAJC6 | Protein coding | -20.0 | -1.8 | -5.1 |
| RALBP1 | Protein coding | -1.0 | -1.7 | -1.1 |
| IRAK3 | Protein coding | -20.0 | -3.8 | -1.9 |
| TSPAN15 | Protein coding | -5.2 | -3.9 | -2.0 |
| TOR1AIP1 | Protein coding | -20.0 | -1.6 | -3.5 |
| HEYL | Protein coding | -20.0 | -1.0 | -1.2 |
| KCNMB3 | Protein coding | -2.3 | -2.7 | -5.2 |
| ABHD17B | Protein coding | -3.0 | -1.8 | -1.8 |

| Gene ID | Function | Inflammation M value | Dysplasia M value | Cancer M value |
|---|---|---|---|---|
| L1TD1 | Protein coding | -4.1 | -1.2 | -1.5 |
| BSDC1 | Protein coding | -1.0 | -1.8 | -1.1 |
| NDC1 | Protein coding | -4.7 | -2.0 | -1.2 |
| GNG12 | Protein coding | -3.0 | -1.2 | -5.0 |
| KLHL8 | Protein coding | -1.4 | -20.0 | -1.9 |
| EPB41L5 | Protein coding | -1.0 | -1.1 | -1.5 |
| HIVEP3 | Protein coding | -1.8 | -2.7 | -2.9 |
| CDH26 | Protein coding | -1.5 | -1.4 | -1.6 |
| NIF3L1 | Protein coding | -2.2 | -2.6 | -4.9 |
| EXO5 | Protein coding | -4.3 | -1.2 | -4.1 |
| SH3D21 | Protein coding | -4.0 | -1.8 | -1.6 |
| ANKRD13C | Protein coding | -6.2 | -1.3 | -4.8 |
| TM2D1 | Protein coding | -4.1 | -1.0 | -4.2 |
| AXDND1 | Protein coding | -5.7 | -1.5 | -20.0 |
| KLHDC7A | Protein coding | -1.2 | -1.2 | -1.4 |
| C1orf87 | Protein coding | -4.0 | -1.2 | -20.0 |
| ZNF280A | Protein coding | -1.5 | -1.2 | -3.4 |
| UBE2U | Protein coding | -5.6 | -1.1 | -2.8 |
| GLIS1 | Protein coding | -3.7 | -2.5 | -1.9 |
| CFAP57 | Protein coding | -1.3 | -20.0 | -2.1 |
| KANK4 | Protein coding | -3.5 | -1.0 | -3.8 |
| ZNF169 | Protein coding | -2.7 | -1.3 | -1.3 |
| SLFNL1 | Protein coding | -4.2 | -1.8 | -1.2 |
| LINC00612 | NCRNA | -1.0 | -1.1 | -3.1 |
| PCSK9 | Protein coding | -1.5 | -1.2 | -2.1 |
| FAM151A | Protein coding | -3.7 | -1.3 | -1.2 |
| CENPW | Protein coding | -4.2 | -4.7 | -3.7 |
| LDLRAD1 | Protein coding | -2.0 | -3.2 | -1.1 |
| C1orf226 | Protein coding | -5.6 | -1.5 | -20.0 |
| LINC00460 | NCRNA | -1.3 | -1.3 | -1.4 |
| PLIN4 | Protein coding | -1.1 | -1.6 | -1.5 |

**Table 13. Hypomethylated promoters common to all stages of CAC**

The 63 genes were found to be hypomethylated compared to the coupled control samples, in inflamed, dysplastic and cancer epithelium but not in the normal epithelium.

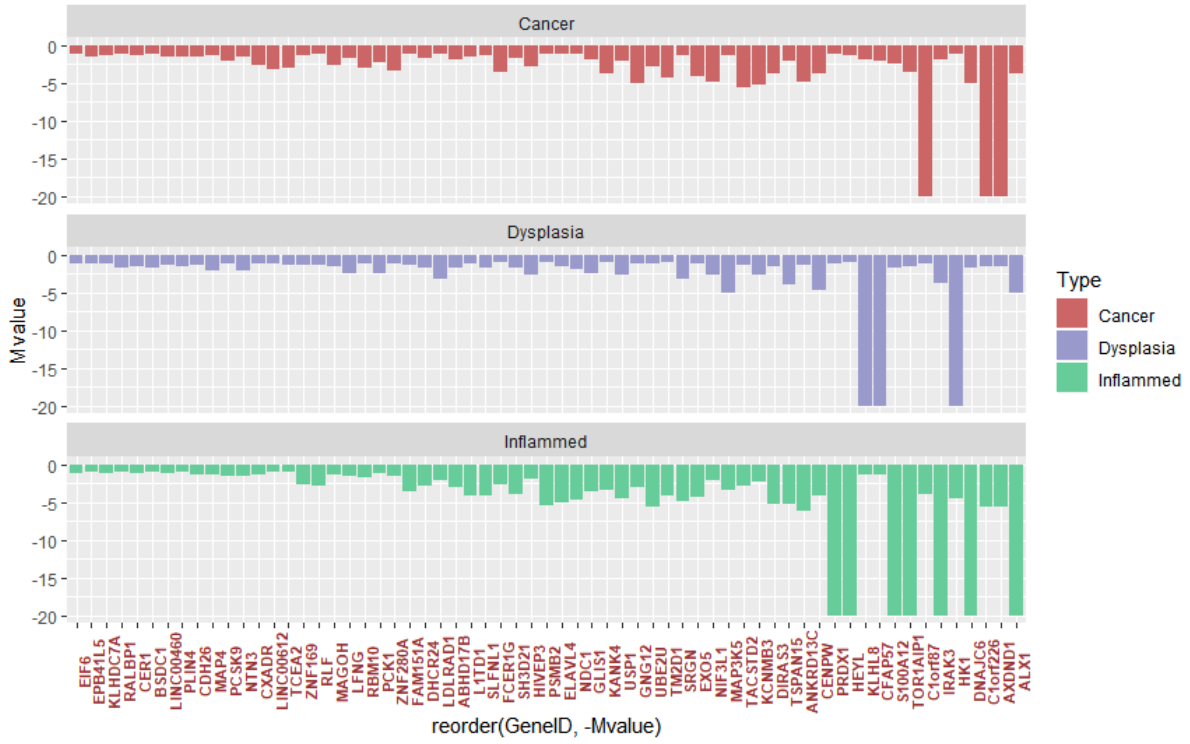



**Figure 21. Hypomethylated promoters in different tissue types**

Hypomethylated promoters in different stages of CAC arranged in order of the mean methylation level.

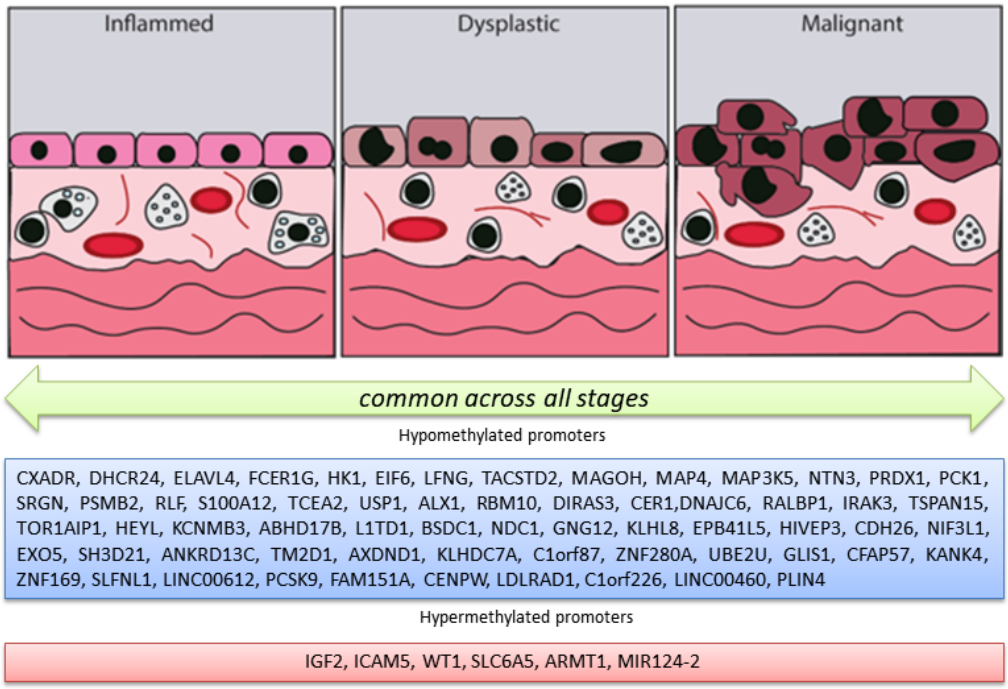



**Figure 22. Common methylation markers across all stages of CAC**

Schematic representation of common methylation markers across all stages of CAC. It is important to recognise those genes involved in the carcinogenic process as some of these gens could be related only to the inflammatory process. The control samples to detect the DMRs were taken from the adjacent tissue to negate this effect.

In my opinion, a useful biomarker in UC needs to predict those at high risk of developing CRC at the stage of inflammation for several reasons. Firstly, the current practice intervenes at the stage of dysplasia and those with dysplasia have been shown to harbour occult cancer at a very high rate (as high as 30%). Secondly, the most effective method of surveillance thus far is serial colonoscopies which are invasive, costly and has a risk of complications such as bowel perforation. Thirdly, patients with long-standing UC who are on immune suppressants and steroids do badly with ileo-anal pouch procedures. Performing restorative surgery once cancer is diagnosed is also still debatable. Therefore, for a biomarker to be useful it needs to

be able to predict the high risk of developing cancer at the early stage where the best outcome from restorative surgery can be offered. Additionally the burden of 'interval cancers' could be addressed with early intervention for high-risk individuals. Interval cancers are those that present between two surveillance colonoscopies and are generally highly aggressive. Within a year or two these cancers tend to develop to be locally advanced in a previously non-neoplastic colon. These may harbour specific high risk molecular characteristics that need to be recognised at an early stage of the disease and patients be offered restorative surgery.

### 4.3.5 Promoter methylations linked to CRC.

Differentially methylated genes may present across all stages due to their relationship to inflammatory process. Methylation changes in genes related to inflammatory changes in UC have been identified (52, 54, 58). Therefore I selected only CRC-associated genes for further analysis.

From the identified hyper and hypomethylated genes we then searched for those genes which have been previously linked to CRC. Seven hypomethylated genes from the 63 and 5 hypermethylated genes from 7 have been linked to CRC with strong evidence (**Table 10**).

| Gene name | Methylation level | | | Function and relation to CRC | Reference |
|---|---|---|---|---|---|
| | INF | DYS | CAC | | |
| TACSTD2 | -2.807 | -1.322 | -5.644 | Transmembrane receptor with $Ca^{2+}$ transporter activity. Acts a s a co-transcriptor for *C-MYC*. Overexpressed in CRC | (198, 199) |
| eIF6 | -1.115 | -1.123 | -1.041 | Transcription initiation factor – found to be over expressed in CRC | (200-202) |
| PRDX1 | -20 | -1.205 | -1.064 | Encodes for an antioxidant enzyme. mRNA and protein levels of PRDX1 shown to be markedly increased in CRC tissues. Identified as a potential biomarker. | (203-205) |
| USP1 | -4.524 | -2.553 | -2.083 | Member of the ubiquitin-specific processing (UBP) family of proteases. Found to be significantly increased in some CRCs. Elevated USP1 level is associated with short overall survival and with advanced stages of cancers. | (206-208) |
| IRAK3 | -20 | -3.807 | -1.874 | Receptor kinase associated with interlukin-1 receptor. High expression is associated with CAC in mouse models. | (209, 210) |
| LFNG | -1.556 | -2.379 | -1.688 | TGFBR2 signalling affects Notch1 glycosylation via regulation of glycosyltransferase LFNG expression. Found to be expressed highly in crypts in CRC. | (211-213) |

| Gene name | Methylation level | | | Function and relation to CRC | Reference |
|---|---|---|---|---|---|
| | INF | DYS | CAC | | |
| S100A12 | -20 | -1.652 | -2.408 | A calcium-, zinc- and copper-binding protein involved in the regulation of inflammatory processes and immune response. Found to be elevated in UC and CRC both. Proposed as a biomarker | (214-216) |
| ICAM5 | 1.485 | 20 | 2.632 | Intercellular adhesion molecule. Gene reported to be methylated in CRC. | (217-219) |
| IGF2 | 1.888 | 1.807 | 1.659 | Methylated in CRC – one of the 5 genes in identified gene panels as biomarkers. When methylated increases *WNT* signal. | (220-222) |
| MIR124-2 | 3.585 | 1.109 | 1 | MIR124 has a suppressive role in CRC. Targets CDK6. Inhibits Warburg effect. Identified as a potential marker in CAC and its down regulated in CRC. Linked to inflammation through STAT3 pathway | (62, 223, 224) |
| WT1 | 2.874 | 4.285 | 2.064 | Tumour suppressor gene. Over expressed in sporadic cancers but methylated in these samples. | (225, 226) |

**Table 14. Genes related to CRC from the differentially methylated genes**

Out of the differentially methylated genes that are common to all three stages, these genes have been already strongly linked to CRC.

Detecting differentially methylated genes that are strongly related to CRC from the early stage of inflammation persistently present through dysplasia and cancer suggest the potential of these genes being biomarkers to predict CAC.

### 4.3.6 Expression level of selected candidate genes

To validate the methylation pattern of the genes we performed RT-qPCR on RNA extracted from the dysplastic lesion and adjacent tissue. With predesigned primers for TACSD2, eIF6, USP1 and LFNG expression levels were measured in relation to POL2A housekeeping gene (**Table 15**).

| **Gene** | CT-test | SD | CT-cont | SD | **ΔΔCT** | 2^**ΔΔCT** |
|---|---|---|---|---|---|---|
| *eIF6* | 31.96 | 0.1214 | 33.79 | 0.0612 | -1.55 | 2.94 |
| *TACSD2* | 29.64 | 0.1635 | 32.64 | 0.1921 | -2.72 | 6.61 |
| *LFNG* | 35.29 | 0.1129 | 36.96 | 0.2506 | -1.67 | 3.18 |
| *USP1* | 30.75 | 0.1785 | 30.58 | 0.1527 | 0.167 | 0.89 |

**Table 15. Expression data for the dysplastic sample**.

Except for USP1 the other 3 hypomethylated genes demonstrated lower CT values indicating a higher expression in the dysplastic tissue in comparison to adjacent tissue.

The expression of *eIF6, TACSD2* and *LFNG* correlated with the methylation profile in the dysplastic epithelium compared to the adjacent tissue. Hypomethylation in the promoter region of *USP1* did not correlated with a significant increase in expression (**Figure 23**).

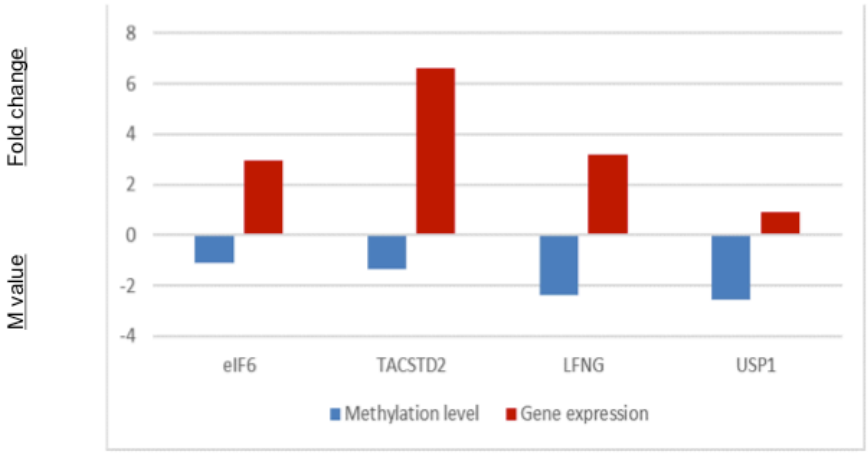



**Figure 23. Gene expression levels correlated to methylation pattern**

Comparison of qPCR data to the promoter methylation levels in the dysplastic colon sample for selected genes. The genes with promoter hypomethylation demonstrate increased level of gene expression. Gene expression levels are presented as 2 ^ΔΔCT and methylation levels in M values for the respective genes. The observed methylation pattern correlates with the expected gene expression.

## 4.4 Analysis of gene body methylation in different stages of CAC

CpG island methylation of the gene body has been correlated with their expression although the significance is not clearly understood. Unlike promoter methylation, the correlation of the gene expression to the gene body methylation is not straightforward. Conflicting patterns of correlation between CpG methylation in the gene body and expression levels of the same gene have been observed (227).

I searched for common genes with gene body methylation across all stages of CAC. Four hundred and eighty three hypo methylated and 14 hypermethylated genes, which were common across all stages and not present in normal colonic mucosa,

were detected. From the hypomethylated genes, the ones with maximum level of methylation level in the cancer epithelium were identified (**Figure 24**).


**Figure 24. G**enes with **a maximum methylation level in the gene body**

The genes were arranged according to the methylation level and those with the highest level of methylation in the cancer were selected. Highest M value of -20 was observed in 92 of the genes out of the 483.

From the identified 92 genes those that have a direct link to CRC carcinogenesis were identified. Ten genes with gene body hypomethylation and 2 genes with gene body hypermethylation met the criteria. Details of the methylation level and the interaction of each gene with CRC are shown below (**Table 16**).

| Gene name | Methylation level | | | Function and relation to CRC | Reference |
|---|---|---|---|---|---|
| | INF | DYS | CAC | | |
| ITGA8 | -1.39 | -1.40 | -20 | Integrin family mediates signal transduction by binding to the extracellular matrix via adhesion receptors. Lower expression has been shown to be lower in CRC and suggested as a diagnostic biomarker. | Gong, Ruan (228), Kok-Sin, Mokhtar (229) |
| ADAM23 | -1.04 | -1.19 | -20 | Family of membrane anchored cell surface proteins involved in Cell adhesion. Promoter hyper methylation is observed in CRC. | Choi, Kim (230), Kalinkova, Zmetakova (231), Takada, Imoto (232) |
| CDH12 | -1.64 | -1.25 | -20 | A membrane-spanning Ca2+-dependent homophilic adhesion receptors. Increases cancer cell migration and invasion via promoting EMT. Also a predictor of poor prognosis in CRC. | Zhao, Li (233), Ma, Zhao (234) |
| ZEB2 | -2.45 | -2.60 | -20 | An epithelial to mesenchymal transition inducing transcription factor. Has a risk stratification value in CRC. High expression is linked to higher invasion and metastasis | Li, Wang (235), Sreekumar, Harris (236), Vu, De Roo (237) |
| RYR3 | -1.82 | -1.44 | -20 | A ryanodine receptor controlling influx of Ca2+ in to the cell. One of the key genes identified as predictors of progression of an adenoma to a CRC. | Lin, Raju (238), Villacis, Miranda (239), Lee, Chun (240) |
| ETS1 | -3.10 | -1.22 | -20 | Is a proto-oncogene and a transcription factor known to control the expression of a number of genes involved in extracellular matrix remodelling. Shown to be expressed highly in CRC and adenomas with high-grade dysplasia. | Gu, Cai (241), Nakayama, Ito (242), Li, Wu (243) |

| Gene name | Methylation level | | | Function and relation to CRC | Reference |
|---|---|---|---|---|---|
| | INF | DYS | CAC | | |
| TMEFF2/ HPP1 | -1.49 | -1.97 | -20 | Transmembrane protein with growth factor like effect. Shown to be methylated in both sporadic CRC and CAC. | Azuara, Rodriguez-Moranta (52), Lin, Taylor (244), Belshaw, Elliott (245), Saito, Kato (246) |
| MYSM1 | -4.89 | -1.42 | -20 | A chromatin-binding transcriptional cofactor. Shown to be highly expressed in CRC compared to adjacent tissue. Predicts poor survival in CRC. | Li, Li (247) |
| USP1 | -6.34 | -1.58 | -20 | Deubiquitylating enzyme that participates in DNA damage response and cellular differentiation pathways. Found to be significantly increased in some CRCs. Elevated USP1 level is associated with short overall survival and with advanced stages of cancers. | Liu, Zhu (206), Ma, Tang (207), Xu, Li (208) |
| ABCA13 | -1.15 | -2.34 | -20 | ATP-binding cassette (ABC) transporters play a crucial role in the development of resistance by the efflux of anticancer agents outside of cancer cells. High expression improves response to chemotherapy. | Hlavata, Mohelnikova-Duchonova (248) |
| WTAS-1 | 2.87 | 4.28 | 2.06 | WT is a tumour suppressor gene. WT1-AS is an antisense that inhibits the expression of WT1 gene. WT1 shows increased expression in CRC. | Oji, Yamamoto (226), Kaneuchi, Sasaki (249), Malik, Wallace (250) |
| CABIN1 | 1.85 | 1.02 | 1.11 | Endogenous inhibitor of calcineurin. Involved in calcium induced cell death and apoptosis. Shown to predict tumour recurrence in CRC. | Watanabe, Kobunai (251), Yi, Kim (252) |

**Table 16. Gene body DMRs annotated to sporadic CRC pathways**

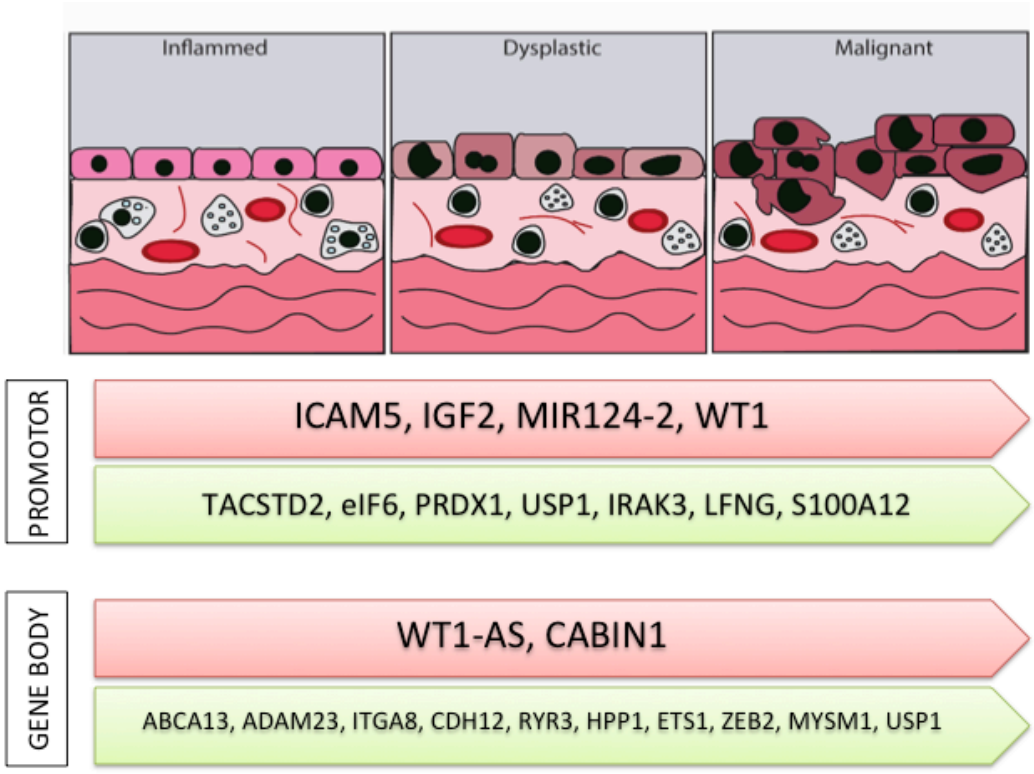



**Figure 25. Potential markers of promoter and gene body methylation**

All genes are differentially methylated and can be detected in inflamed, dysplastic and cancer tissue but not in the normal colon epithelium. All genes have been linked to CRC development.

In total 23 potential differentially methylated markers were identified from the WGBS data (**Figure 25**). Eleven were promoters (4 hypermethylated and 7 hypomethylated) and 12 were gene methylations (2 hypermethylated and 10 hypomethylated). All these potential markers have strong links to the development of propagation of CRC.

## 4.5 Conclusion

UC is an idiopathic chronic inflammatory condition predisposing the affected to CAC. Surgical removal of the entire colon and replacing with a small bowel pouch is

recommended for neoplastic changes due to longstanding UC. However detection of neoplastic changes is currently done with surveillance colonoscopy and there is evidence that some patients have already developed cancer by the time they receive surgery. Therefore a predictive biomarker, which could predict neoplastic changes, will benefit the patients by facilitating early decision making to surgically resect the colon prior to cancerous changes. Outcome from surgery has been shown to be significantly better when performed prior to carcinomatous changes.

Although the chronic inflammation was traditionally thought to be the reason behind malignant transformation, the focus is shifting towards genetic changes, which occur early in the process. It has been demonstrated that CRC-related genetic mutations are prevalent in the inflamed colonic mucosa even before neoplastic changes. However persistent genetic mutations have not been identified thus far, redirecting the focus of the investigators to epigenetics.

This results chapter discusses the outcome of WGBS done on different stages of UC associated CRC. The methylation changes were matched with control unaffected mucosa or the blood from the same patients. Epithelial cells from a normal (non-UC) colon were also sequenced (and matched with blood from the same individual), which was used to differentiate exclusively UC related methylation changes.

A limitation of this study is the fact that it utilises only 8 samples, with one test and one control sample from each tissue type. However, I have used laser-dissected epithelial cells from each sample to prevent the contamination from stromal cells, which have a different genetic composition from epithelial cells. Also I have carried out WGBS on a next-generation sequencing platform, which has not been done on purified epithelial cells before. Performing WGBS on isolated epithelial cells from both test and control samples is both labour and resource-intensive. The main

objective was to demonstrate the feasibility of such a design with highly specific (for epithelial cells) deep sequencing. Although the number of samples used is low, the specificity of the data is higher. Earlier studies had analysed epithelial cells isolated with EDTA and used a targeted sequencing approach. The objective of this study was to perform a comprehensive analysis of the methylation pattern in epithelial cells which are the source of CAC. Future studies are need to utilise this design and include a larger number of samples, given the availability of resources to obtain robust specific data on the promoter methylation pattern in CAC. Another area of concern in analysing methylation changes is the selection of an appropriate control sample. As some methylation changes may be inherited and some can occur with the normal aging process, it is pertinent to identify the disease specific changes by comparing with uninvolved tissue samples. For the current study, the best available control samples were selected for comparison of methylation changes as age-related and inherited methylation had to be subtracted. For those with neoplastic changes, the adjacent non neoplastic tissue was selected, and for those with inflammation, the buffy coat was selected. Since the inflamed colons, which were subjected to surgery, had pan colitis, the buffy coat had to be used as the control sample for methylation changes. The same approach was used for the healthy colon sample, which had to be tested against the best available control sample. The methylation in the buffy coat reflects the methylation changes in whole blood (253) and has been used to accurately demonstrate the age related methylation changes (254, 255). It has been demonstrated that buffy coat does not demonstrate correlating methylation changes of cancer related genes observed in primary malignant tumours (256, 257). Common persistent methylation changes present from the stages of inflammation along dysplasia and cancer were detected. In general more hypomethylated genes

(compared to matched controls) were detected in cancer tissue. Gene annotation analysis revealed that the top 3 pathways annotated to the differentially methylated genes in all three stages were metabolic pathways, pathways in cancer and PI3K pathway.

I have identified promoter hypomethylation in 63 and promoter hypermethylation in 6 genes that were common to all stages of CAC and were not present in normal colonic epithelium. Out of these the genes that were associated with CRC based on literature were separated. Seven (*TACSD2, eIF6, USP1, PRDX1, S100A12, IRAK3, LFNG*) out of the 63 hypomethylated promoters and 4 (*ICAM5, IGF2, WT-1, MIR124-2*) out of the 6 hypermethylated promoters were in genes strongly associated with CRC. Although a limited number of genes were included, gene expression data on selected genes confirmed the sequencing data. The higher expression levels in eIF2, TACSD2 and LFNG with promoter hypomethylation validate the sequencing data to an extent. There are a large number of differentially methylated genes that are identified yet to be validated with expression data. Although it is widely accepted that promoter methylation and gene expression has an inverse relationship, this general rule has been challenged, particularly in cancer (258). The correlation between methylation and gene expression has been shown to be applicable to particular genes such as transcription factors more frequently than to other genes (258). In the limited number of genes that have been tested in the study, I have included a membrane protein (LFNG), a membrane transporter (TACSTD2), an enzyme involved in DNA repair (USP1) and a transcription factor (eIF2) to have a representation of commonly involved mechanisms in carcinogenic transformation. A future study with less resource and time limitations, I envisage assessing the expression level of all 63 genes that were identified as common to all stages in CAC.

Promoter hypo methylation of *TACSD2, LFNG* and eIF6 are linked to *C-MYC, NOTCH* signalling and transcription initiation with well-established links to CRC. High expression of these genes from the inflammatory stage may indicate their important role and predictive value in CAC. Also *S100A12* and *IRAK3* are both highly expressed in UC and CRC. Products of both genes have been shown to be elevated in UC and suggested as a biomarkers previously.

Influence of gene body methylation on gene expression is less well studied compared to promoter methylation (259). Gene body methylation downstream of the promoter region has shown to influence gene expression (260). There were 423 hypomethylated genes and 14 hypermethylated genes that were common to inflamed, dysplastic and cancer epithelial cells. Out of the total 10 hypomethylated genes and 2 hypermethylated genes were previously linked strongly to CRC. *USP1*, which demonstrated hypomethylated promoter region, also had gene body hypomethylated in all stages. Increased expression of *USP1* has shown to be associated with poor clinical outcome in sporadic CRC which makes the finding more significant.

This is the first study, which has looked in to cancer related epigenetic markers common to all crucial stages (inflammation, dysplasia and cancer) of CAC development. In doing so we intended to identify epigenetic markers that can predict the development of cancer early in the disease process. Identifying genes that are involved in growth, differentiation and influence invasiveness of cells, with promoter hypo methylation in the early noncancerous stages, may indicate that the epigenetic changes towards CAC development is an early phenomenon. Also these data indicate a CAC pathway that is independent of the chronic inflammation.

These markers need to be validated in a large prospective cohort of patients with UC before they can be implemented in a clinical setting. Markers that can be identified at early inflammatory stage is important since they could guide the clinician in making an early decision to offer restorative surgery which prevents the development of cancer an ensure improved outcome from surgery.

# 5.0 Results: 2

## 5.1 Analysis of methylation pattern in already recognised markers

### 5.1.1 Scope of the chapter

As described in the introduction, the sequence of genetic mutations leading to CAC varies from sporadic CRC. Mutations in genes such as *KRAS, P53* and *FAP* are common for both forms of cancer. However epigenetic changes are thought to play an early role in CAC (261-263). Key driver genes such as APC, KRAS, BRAF, PIK3CA, SMAD4 and p53 that are identified in sporadic CRC have not been identified to play a major role in CAC. Evidence on the epigenetic modifications of those key genes in CAC is lacking. The possibility of some of the key driver genes being silenced or activated through methylation is explored in this section. Existing evidence suggests several methylated genes to have potential as markers in CAC(51, 52, 57). In this chapter, an analysis of the methylation pattern of eleven genes from this group was performed. Other investigators have also suggested few methylated genes as stool based biomarkers in sporadic CRC (245, 264). The sequencing data is analysed to assess the presence of such methylation markers in the current study samples.

### 5.1.2 Chapter summary

Out of the recognised 6 key genes of CRC, 5 genes show differential methylation in different stages of CAC in the current samples. There is no consistent pattern of methylation that allows them to be considered as predictive markers. Out of the proposed markers of CAC, *SFRP2* and *P16* demonstrate a consistent pattern of hypo- and hyper-methylation of the gene bodies respectively. Previously identified tool based marker, *SEPTIN9* shows a consistent hypomethylation in inflammation and dysplasia, which is absent in the normal colon making it a potential marker.

### 5.1.3 Methylation of key driver genes in sporadic CRC

Six key driver genes in sporadic CRC that were previously described were analysed for methylation changes (265, 266). Out of the 6 genes *FAP*, *KRAS, BRAF, PIK3CA, SMAD4* and *p53,* only *PIK3CA* showed differential methylation in the CAC epithelium. Both P53 and BRAF had gene body methylation in inflamed sample while *BRAF* demonstrated promoter methylation. *KRAS* and *SMAD4* had promoter hypomethylation in the normal sample (**Table 13**).

| Gene | **NORMAL** | | **INFLAMED** | | **DYSPLASIA** | | **CANCER** | |
|---|---|---|---|---|---|---|---|---|
| | Promoter | Body | Promoter | Body | Promoter | Body | Promoter | Body |
| KRAS | -1.16 | *ND* | *ND* | *ND* | *ND* | *ND* | *ND* | *ND* |
| BRAF | *ND* | *ND* | -3.94 | -20 | *ND* | *ND* | *ND* | *ND* |
| PIK3CA | *ND* | *ND* | *ND* | *ND* | *ND* | *ND* | *ND* | -20 |
| SMAD4 | *ND* | -1.06 | *ND* | *ND* | *ND* | *ND* | *ND* | *ND* |
| p53 | *ND* | *ND* | *ND* | -1.08 | *ND* | *ND* | *ND* | *ND* |

**Table 17. Promoter and body methylation pattern of key driver genes in CRC**

The key driver genes identified in sporadic CRC did not demonstrate a consistent methylation pattern in the study samples.

*ND – not detected*

None of the key driver genes showed differential methylation in the dysplastic tissue compared to the adjacent colonic mucosa. Also, only *PIK3CA* showed a significantly higher level of hypomethylation in cancer tissue compared to the adjacent non cancer tissue. This may be due to the fact that the key genes are not differentially methylated in the process of CAC or the field changes have caused methylation changes in adjacent tissues as well. *BRAF* and *p53*, are hypomethylated in inflamed tissue compared to buffy coat.

## 5.2 Differential methylation of previously described key genes of CAC

I analysed the methylation pattern of key genes that were observed to be differentially methylated by other investigators in CAC (**Figure 26**). *RUNX3, MINT1, MYOD1, APC2, P16, SFRP2, SFRP4, SFRP5, EYA4, ESR* and *WIF1* has been found to be differentially methylated in colitis associated cancer by several authors (57, 246, 267, 268). *SFRP2* gene body hypomethylation and *P16* promoter hypermethylation were the only two DMRs that could be seen across all stages of CAC and not in normal epithelium (**Figure 27**).

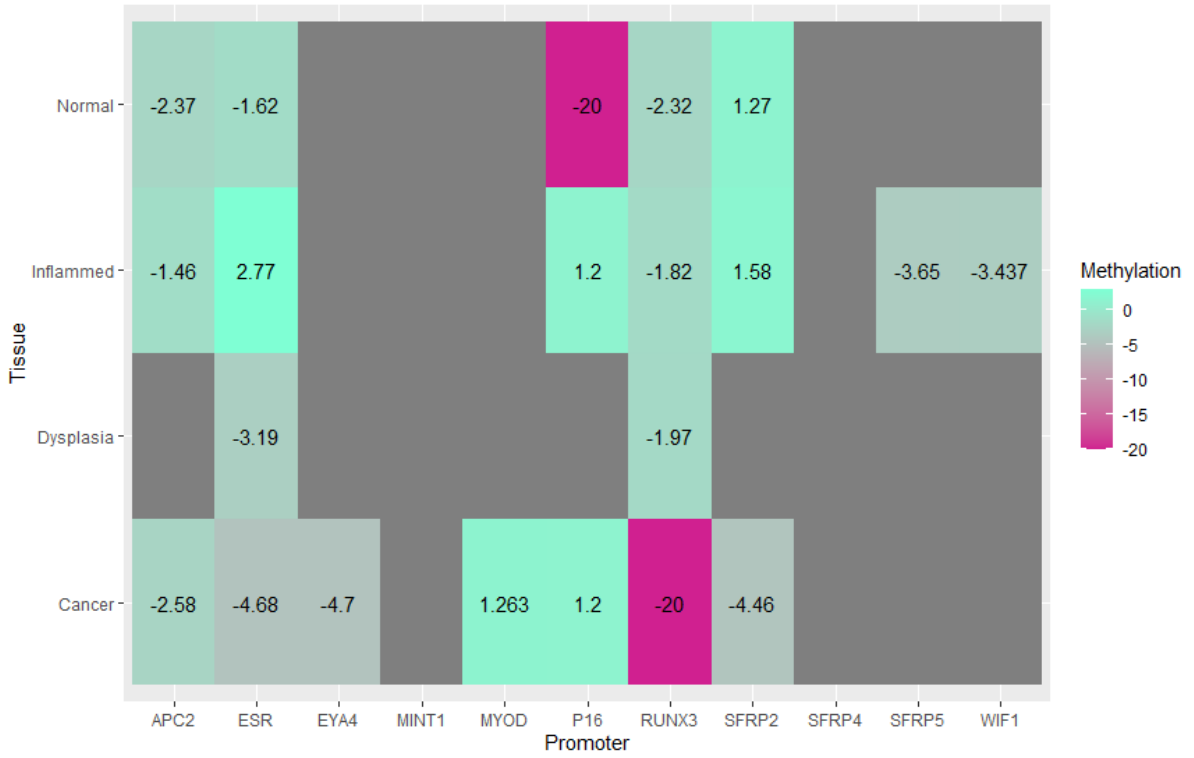



**Figure 26. Heat map of the promoter methylation in key driver genes.**

The methylation levels (M values) of the key genes identified key driver genes in sporadic CRC were compared across different stages of CAC and in the normal colonic epithelial sample. Promoter of P16 shows a shift from being hypomethylated in normal epithelium to being hypermethylated in the inflamed and cancer epitheliums. SFRP2 has shifted from being hypermethylated in both normal and inflamed samples to being hypomethylated in cancer. Gray areas indicate the promoter that do not shoe DMRs compared to the control samples.


**Figure 27. Gene body methylation pattern of key driver genes.**

Most genes do not show a distinct pattern of gene body methylation pattern compared to the normal epithelium, that Is consistent along different tissue types. SFRP2 gene body shows similar pattern to its promoter region with shifting from being hypermethylated in normal epithelium to being hypomethylated consistently in the inflamed, dysplastic and cancer epitheliums.

## 5.3 Methylation pattern of previously recognised stool biomarkers

Several biomarkers in stool have been suggested to predict CAC (214, 264, 269, 270). Methylation of NDGR4, VIM and SEPT9 has been repeatedly shown to be methylation biomarkers in stool for sporadic CRC and CAC (264, 269-271). However different studies have demonstrated varying sensitivities of the marker as a predictor. In the current study promoter hypomethylation of SEPT9 has been shown to have a

distinct pattern in inflamed and dysplastic tissue compared to normal tissue (**Figure 28**).

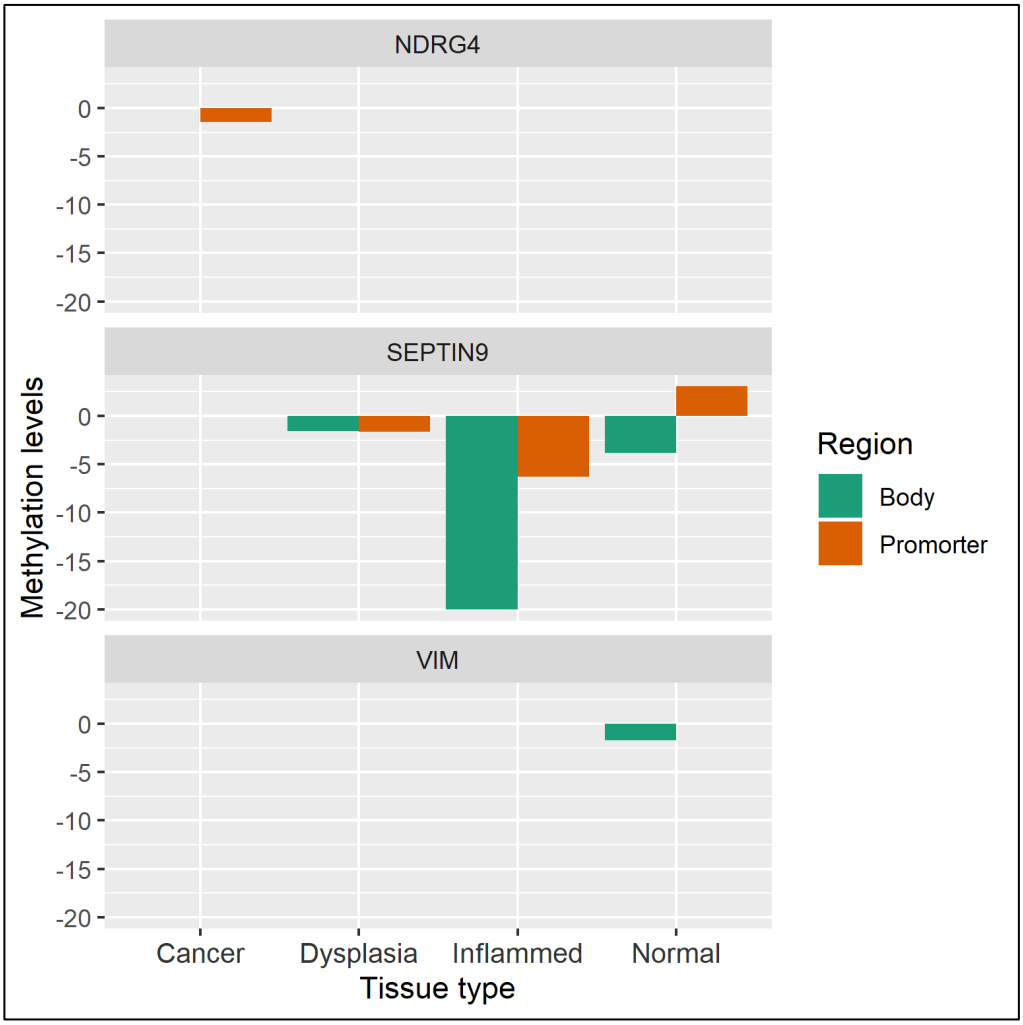



**Figure 28. Differential methylation of potential stool biomarkers**

Methylation pattern of the NDRG4, SEPTIN9 and VIM, which have been suggested as stool based biomarkers of CRC. SEPTIN9 promoter hypomethylation is seen in inflamed and dysplastic tissue in contrast to the normal epithelium. NDRG4 and VIM do not show a consistent methylation pattern.

## 5.4 Gene Ontology (GO) analysis of the suggested candidate genes

### 5.4.1 Gene annotation to biological processes

Using the ToppGene Suit gene enrichment analysis tool, functional annotations analysis was done for the suggested candidate genes (196).

The 27 candidate genes were annotated to two molecular functions; cell adhesion binding and integrin binding (**Figure 25**).

### 5.4.2 Gene annotation to disease processes

The candidate genes annotated to disease processes revealed their significant relations to several disease processes of interest. UC, hyperplastic polyp, adenoma of the large intestine, epithelial carcinomas and sarcomas were amongst them (**Figure 29 and 30**). This increases the relevance of the selected genes to be investigated as an early bio marker of CAC.

Analysing the 15 hypomethylated genes out of the 27 candidate genes the molecular function annotations were made to cell adhesion molecular binding, cadherin binding and MAP kinase activity. These three functions are closely related to cancer cell functions. These GO findings further support the further evaluation of the suggested candidate genes as potential biomarkers of CAC.

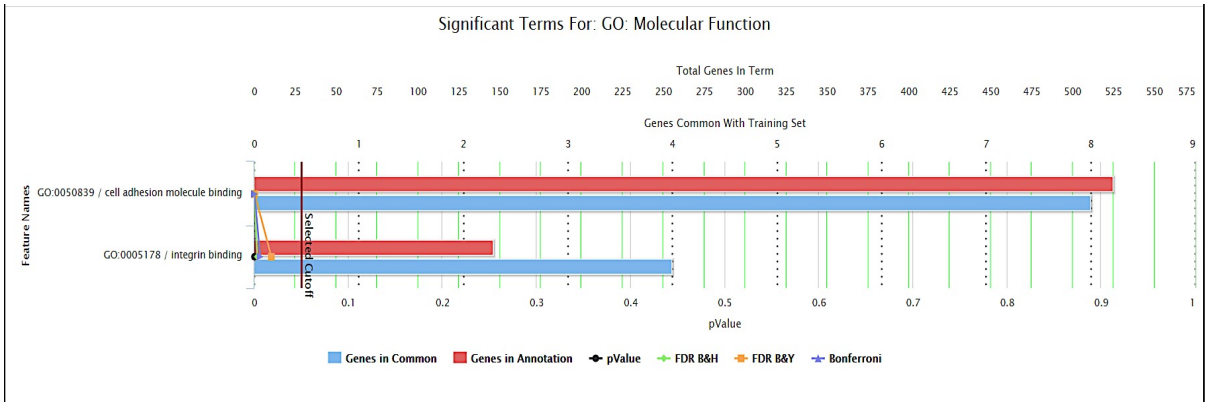



**Figure 29. Molecular function annotation for candidate genes**

The 27 candidate genes with differential methylation patter identified from WGBS based on their persistency and distinction from normal tissue were analysed with the genen enrichment analysis software ToppGene Suit. The gene set was annotated to two molecular functions related to cell adhesion function with the Bonferroni correction and at a significance level of 0.05.

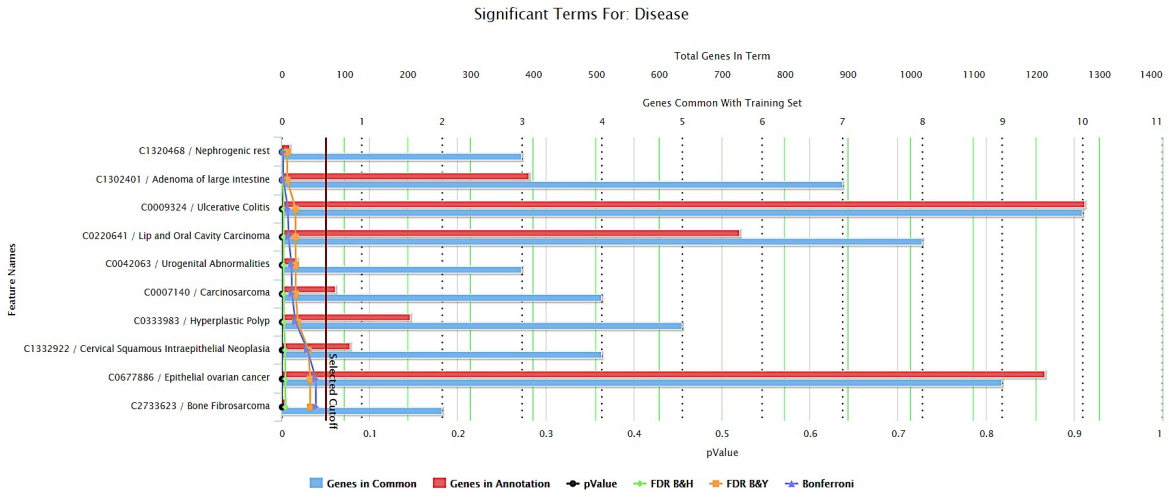



**Figure 30. Gene annotation to disease processes**

The 27 candidate genes were annotated to disease processes using the ToppGene Suit tool with Bonferroni correction and a significance level of 0.05. The genes were annotated to UC, large bowel adenoma and several epithelial carcinomas.

## 5.5 Conclusion

Key driver genes are genes containing a higher mutational burden than expected compared to the background mutation frequency in relation to any cancer. There are 6 key driver genes identified in CRC; *FAP, KRAS, BRAF, P53, PIK3CA* and *SMAD4*. I investigated whether these driver genes demonstrated consistent methylation changes. In the samples studied, differential methylation in these genes was rare compared to the controls. Promoter hypomethylation of BRAF and P53, was detected exclusively during inflammation. Hypomethylation of PIK3CA was detected during cancer but not in the pre-malignant stages, reducing its value as a predictive biomarker. I also analysed the genes that have been frequently recognised as methylated in CAC. Eleven genes that included, *RUNX3, MINT1, MYOD, APC2, P16, SFRP2, SFRP4, SFRP5, EYA4, ESR* and *WIF1* were searched for the pattern of differential methylation in both promoter and gene body. This set of genes have been repeatedly detected to be methylated in CAC although their status in premalignant stages haven't been demonstrated. The differential methylation pattern of these genes was compared with the normal colonic epithelium.

Promoter hypomethylation of *P16* was detected during inflammation and cancer while hypermethylation was detected in the normal epithelium. *P16* is a cell cycle regulator gene which is strongly implicated in malignancies including CRC (272). The observed methylation pattern strongly indicates its potential as a predictive biomarker. *P16* has been demonstrated previously to be overexpressed in the epithelium of UC affected colon (273). Similarly, body hypomethylation of SFRP2, a modulator of the WNT pathway, was detected during inflammation, dysplasia and cancer, in contrast with the hypermethylation observed in the normal epithelium. SFRP2 has been shown to be hypermethylated in CRC as well as expressed highly

in invasive cancers (274, 275) where it has been suggested as a stool based biomarker as well as its expression as a poor prognostic marker in CRC (274). Increased level of hypomethylation in the gene bodies of WIF1, RUNX3 and MINT1 were detected in all stages, although the changes did not distinguish them from the normal epithelium. The products of these genes may have a biomarker potential if they demonstrate higher concentration in blood or stools.

Finally previously suggested stool-based biomarker genes NDRG4, SEPTIN9 and VIM were also analysed. Only SPETIN9 promoter hypomethylation was detected in the inflamed and dysplastic tissue. With its strong links to sporadic CRC, SEPTIN9 is worth being included in a candidate gene panel to be tested as a biomarker. Form these observations P16, SFRP2 and SEPTIN9 needs to be included in a candidate gene panel for predicting CAC that will be evaluated in a prospective cohort of patients.

The GO data from the identified potential biomarkers demonstrates an interesting perspective. Although most of the genes common to all stages of CAC are not previously described as biomarkers the GO enrichment analysis provides evidence to their relationship with both UC and cancer related molecular functions. The genes are significantly related to cell adhesion and MAP kinase activity both are which deranged in cancer cells. The disease annotation to the candidate genes demonstrated their relation to UC, colonic polyps and mucosal cancers in several organs. These findings support the potential of the candidate gene set to be further evaluated as a biomarker panel.

# 6. Results: 3

## 6.1 Expression of SMAD7 in different stages of inflammation associated cancer

### 6.1.1 Scope of the chapter

Expression of SMAD7 has been shown to increase in tissues affected with UC. The increase of this inhibitory molecule is postulated to negate the anti-inflammatory effect of TGFβ1. Although the increase in SMAD7 has been demonstrated during the inflammatory phase, (126, 140-142) data on its behaviour during the progression to colitis-associated neoplasia has not been demonstrated (125, 145, 146, 276-278). In this chapter the changes in SMAD7 protein in the epithelia lcells of the colonic mucosa in different stages of CAC is demonstrated. Also, data exist to demonstrate that the increase in SMAD7 during inflammation is not due to an increase in the synthesis of mRNA, but whether the same remains during the progression to neoplasia has not been explored. One of the aims of this study was to explore the changes in SMAD7 during different stages of CAC and to demonstrate how its transcription is affected.

Further, SMAD7 protein has shown to affect the methylation of three genes, products of which are involved in cell migration and epithelial mesenchymal transition, in breast cancer cells (133). I analysed the sequencing data to explore the possibility of these genes demonstrating a correlation with the biphasic pattern of SMAD7 expression.

### 6.1.2 Chapter summary

This chapter describes the changes of SMAD7 in the colonic epithelium in UC and different stages of CAC. Using Immuno histochemistry (IHC) I demonstrate the biphasic changes in SMAD7 molecule in the colonic mucosa of tissue samples form patients with UC. During the phase of inflammation there is a significant increase in the concentration of SMAD7 compared to on-inflammed tissue. The dysplastic tissue demonstrates a reduction in the expression while a significant rise is seen again in the CAC tissues. This biphasic change in SMAD7 has not been demonstrated before. With IHC for pSMAD3 it is demonstrated that a complementary reduction in this second messenger molecule is not seen during the high concentration of SMAD7. This may indicate that the TGFB pathway escapes the inhibitory action of SMAD7 in UC affected epithelium. Using in situ hybridisation I have demonstrated that the mRNA of SMAD7 does not show an increase during the inflamed or cancer stages. This may suggest alternative mechanisms of increasing SMAD7 concentration during these phases, except for an increase in transcription levels. Analysis of the WGBS sequencing data demonstrated that the methylation of the genes related to EMT (CDH1, CGN and CLDN4) postulated to be affected by SMAD7 does not show a consistent methylation pattern correlating to changes in SMAD7 in UC affected colon.

### 6.1.3 SMAD7 antibody optimisation

After optimisation on sections of human liver cancer, a final concentration of 0.03 µg/µl was selected for the antibody recognising SMAD7 (Figure 27), as expression of the protein in the cytoplasm could be detected with minimal background staining. The isotype primary antibody repeatedly gave negative staining.

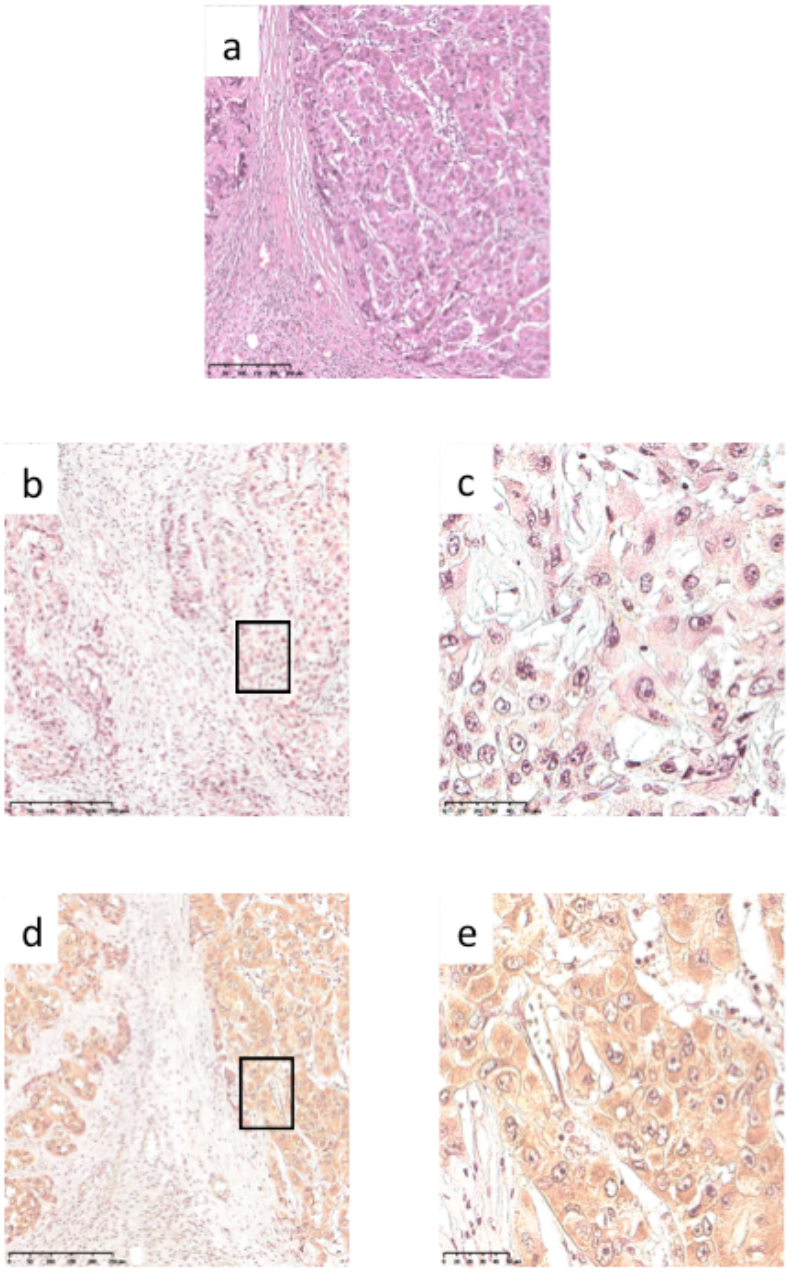



**Figure 27. Optimisation of anti SMAD7 antibody**

a) Haematoxylin and eosin staining of liver cancer tissue section. b) 10X – scale bar - 250µm & c) – 40X – scale bar – 50 µm; Mouse isotype control (0.033µg/µL) with anti-mouse secondary antibody (1:500). d) 10X – scale bar - 250µm & e) 40X – scale bar – 50µm; SMAD7 antibody (0.033µg/µL) with mouse secondary antibodies (1:500).

### 6.1.4 SMAD7 immunohistochemistry on different stages of CAC

Immunohistochemical staining was performed in 53 paraffin fixed samples from 25 patients, to assess the expression of cytoplasmic SMAD7 in non-inflamed (n=12), inflamed (n=12), dysplastic (n=12) and colitis associated cancer specimens (n=17). With image analysis software to assess staining intensity, expression of SMAD7 in the cytoplasm showed a biphasic pattern along the different stages of colitis associated carcinogenesis (**Figures 28 to 31**). Kruskal-Wallis test was used to assess intergroup variability in the population and Mann-Whitney U test was used to compare between individual stages. SMAD7 expression was significantly high in the inflamed compared to the non-inflamed epithelium ($p<0.01$). The expression of the protein is lower in the dysplastic epithelium and showed a wider variation amongst tissue sections. Cancer epithelium demonstrated a significantly higher staining intensity ($p<0.01$) for SMAD7 compared to the dysplastic tissue (**Figure 32**).

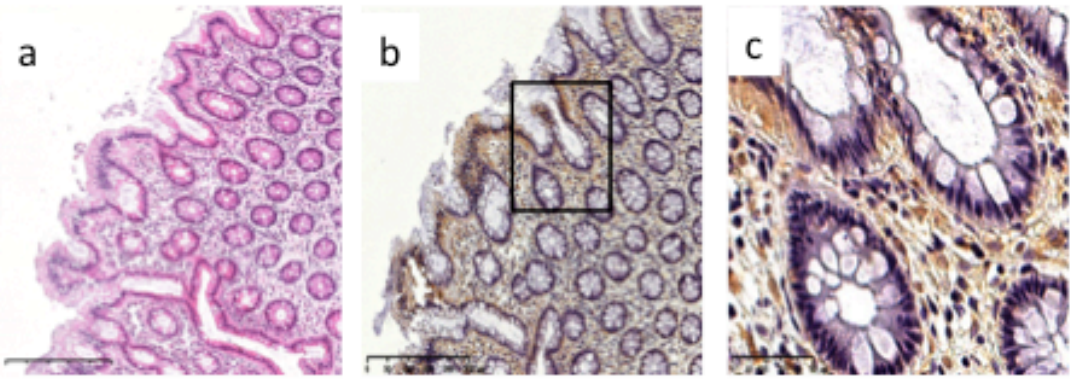



**Figure 28. Expression of SMAD7 in non-inflamed/ non-neoplastic colon**

a) H&E sections of non-inflamed/non-neoplastic colonic mucosa from a patient with UC. b) Serial section of the same tissue stained with anti SMAD7 anti bodies, Scale bar – 250 µm. c) Magnified view showing non-stained cytoplasm of the epithelial cells, scale bar - 50µm.

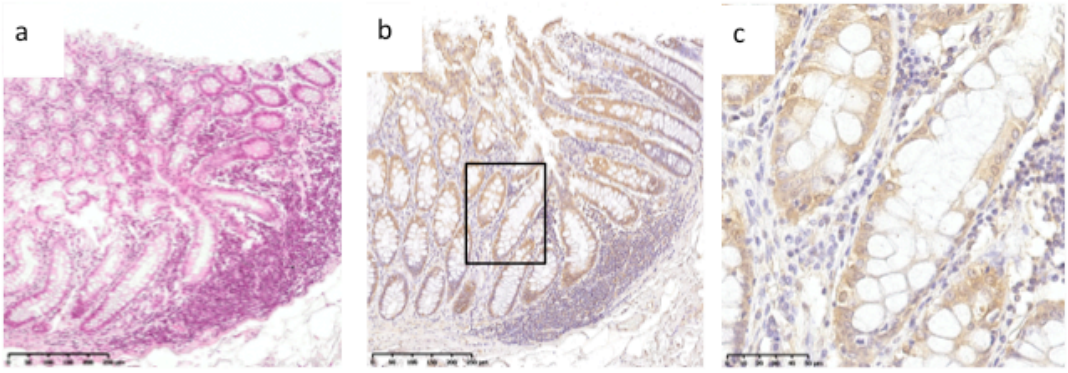



**Figure 29. Expression of SMAD7 in the inflamed colonic epithelium**

a) H&E stained sections of inflamed colon mucosa from patients with UC. b) Serial section stained with anti SMAD7 antibodies (scale bar - 250µm). c) 40X Magnified view of the selection demonstrating dark staining of the cytoplasm of epithelial cells indicating high expression of MSAD7 (scale bar - 50µm).

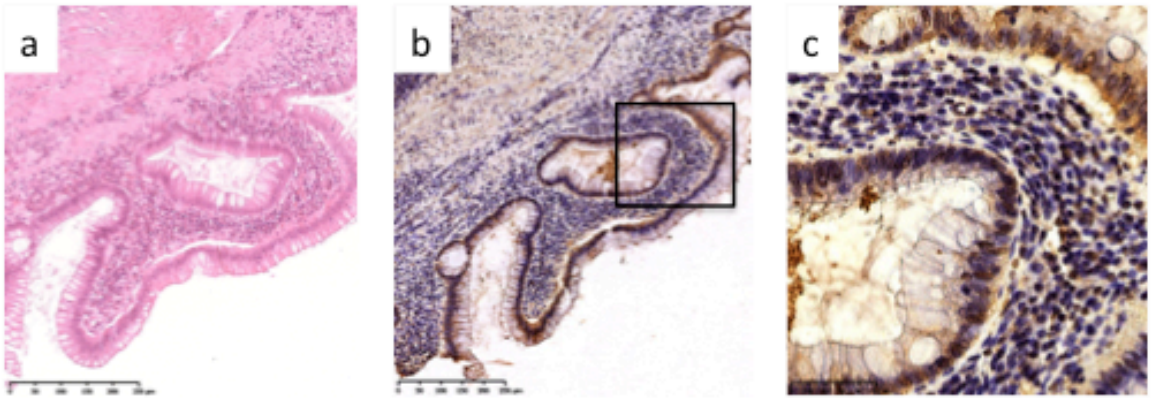



**Figure 30. SMAD7 in dysplastic colonic epithelium in patients with UC**

a) H&E stained sections of dysplastic colon mucosa from patients with UC. b). Serial section stained with anti SMAD7 antibodies; (scale bar - 250μm). c) 40X Magnified view of the selection demonstrating light staining of the cytoplasm of epithelial cells indicating a lower expression of MSAD7 compared to the inflamed tissues (scale bar - 50μm).

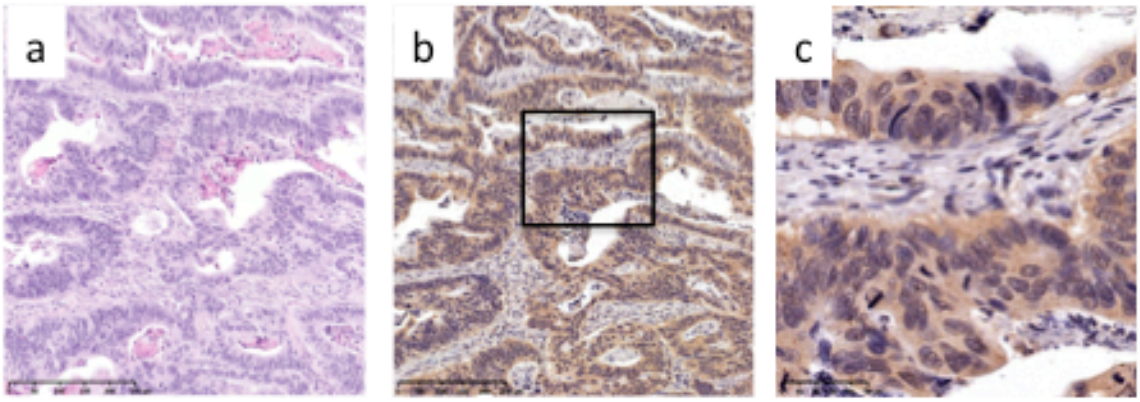



**Figure 31. SMAD7 in the colonic epithelium in patients with CAC**

a) H&E stained sections of CAC from patients with UC. b) Serial section stained with anti SMAD7 antibodies (scale bar - 250μm). c) 40X Magnified view of the selection demonstrating dark staining of the cytoplasm of epithelial cells indicating high expression of MSAD7 (scale bar - 50μm).

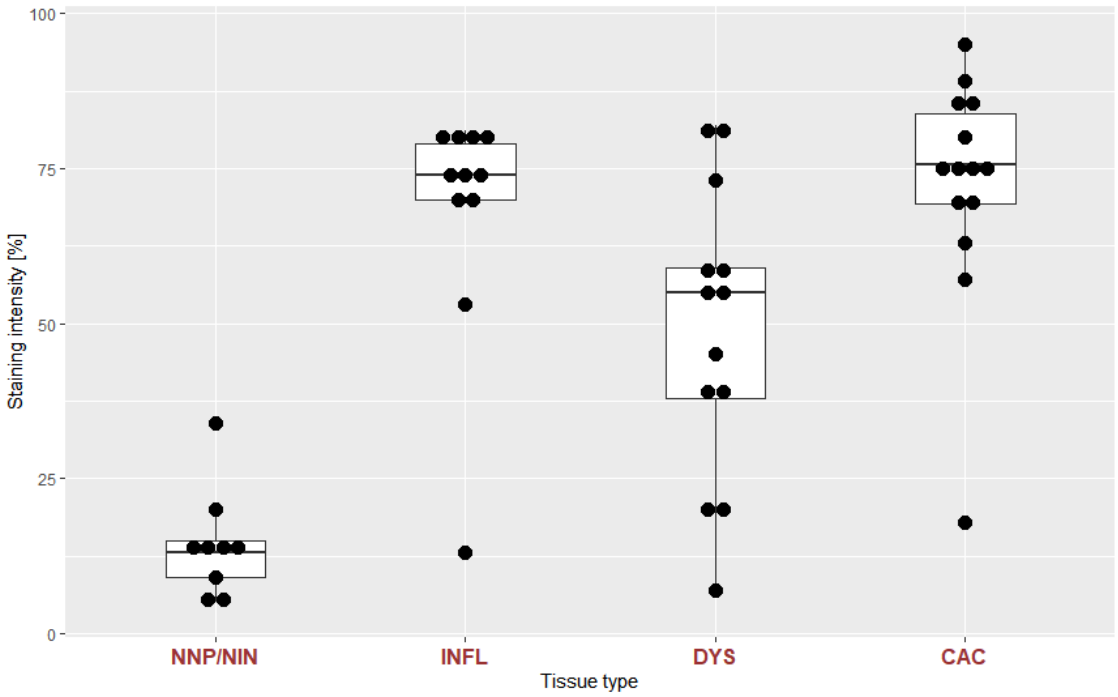



**Figure 32. Expression of SMAD7 in the colonic epithelium**

The expression of SMAD7 protein in non-neoplastic/ non-inflamed (NNP/NIN), inflamed (INF), dysplastic (DYS) and colitis associated cancer (CAC) tissue (n=53). Quantification was done using the image analysing software Visiopharm. Using Kruskal-Wallis test for inter group variability ($p<0.001$) and Mann-Whitney U test to compare between different stages, a significant difference in the expression of SMAD7 in the epithelium of different tissue sections was observed (p values: NNP/NIN vs INF; <0.01, INF VS DYS = 0.01, DYS vs CAC <0.01). The expression pattern was biphasic with high levels in the inflamed and cancer tissue. A significantly low average expression was observed in the dysplastic tissue.

### 6.1.5 Optimisation of pSMAD3 antibody

The working dilution of 1:50 was chosen for pSMAD3 antibody as at this dilution the maximum nuclear staining with minimum background was observed (**Figure 33**).

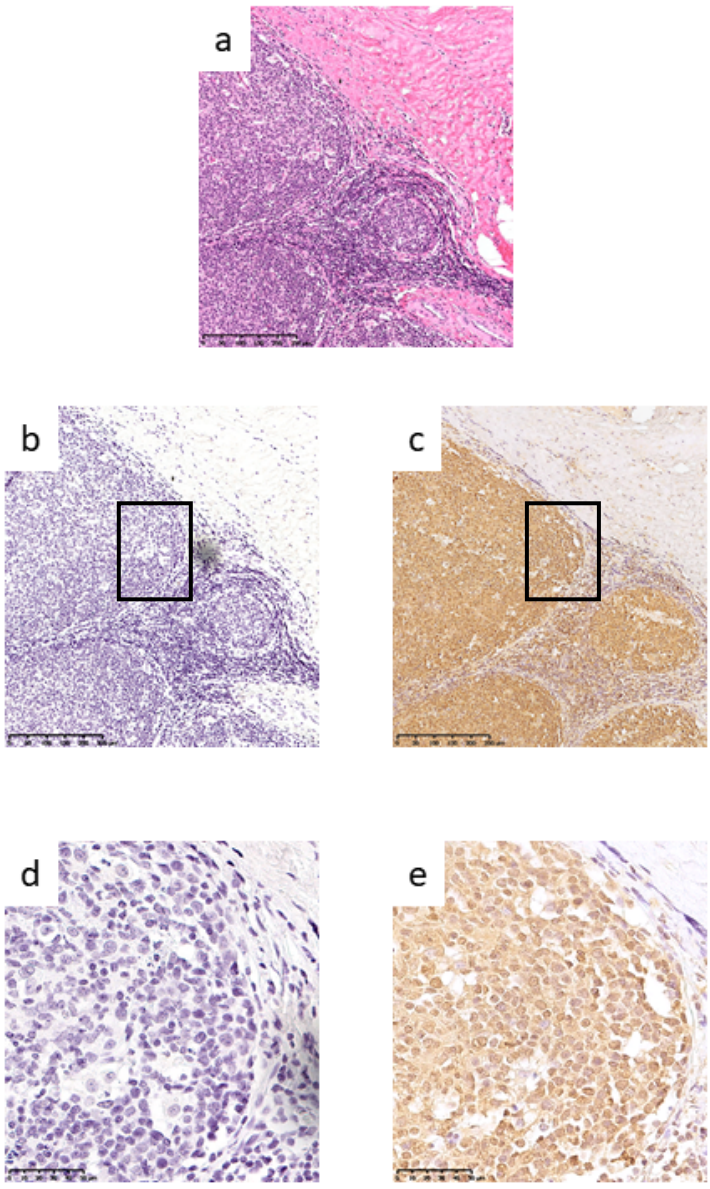



**Figure 33. Anti pSMAD3 antibody optimisation**

a) Haematoxylin and eosin staining of a cross section of tonsil. b) negative control with iso type rabbit antibodies. c) tonsil section stained with anti pSMAD3 antibody

demonstrating dark staining, scalebar - 250µm. d) & e) – magnified views of selected regions from b)&c), scale bar - 250µm

### 6.1.6 Immunohistochemistry for pSMAD3 in different stages of CAC

According to the stipulated TGFB pathway, an increase in SMAD7 is expected to inhibit pSMAD3 in the nucleus due to the inhibition of its phosphorylation. In this cohort of colon samples, the expression of pSMAD3 in the nucleus is not inversely correlated to SMAD7. Nuclear expression of pSMAD3 in the normal, inflamed, dysplastic and malignant epithelial cells do not differ significantly (p=0.17) (**Figure 34 and 35**).

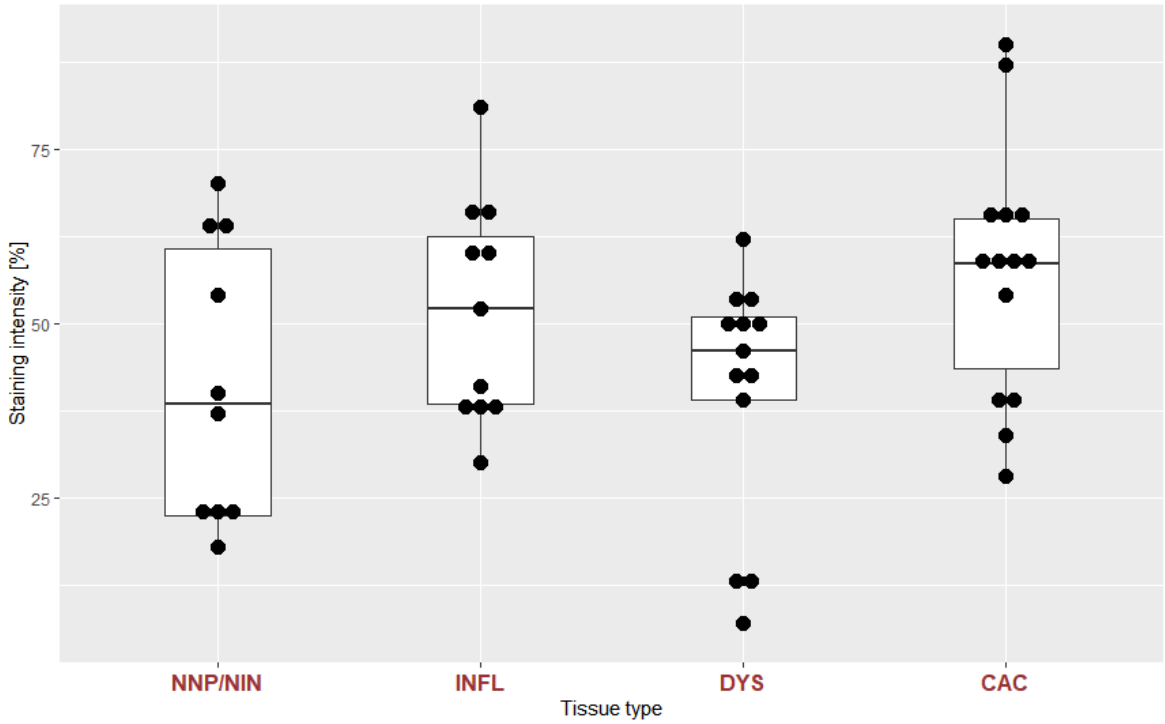



**Figure 34. Expression of pSMAD3 in the colonic epithelium**

The presence of pSMAD3 protein in non-neoplastic/ non-inflamed (NNP/NIN), inflamed (INF), dysplastic (DYS) and colitis associated cancer (CAC) tissue. Quantification was done using image analysing software. There was no significant

difference in the expression of nuclear pSMAD3 in the epithelial cells of different tissue sections (p=0.4).

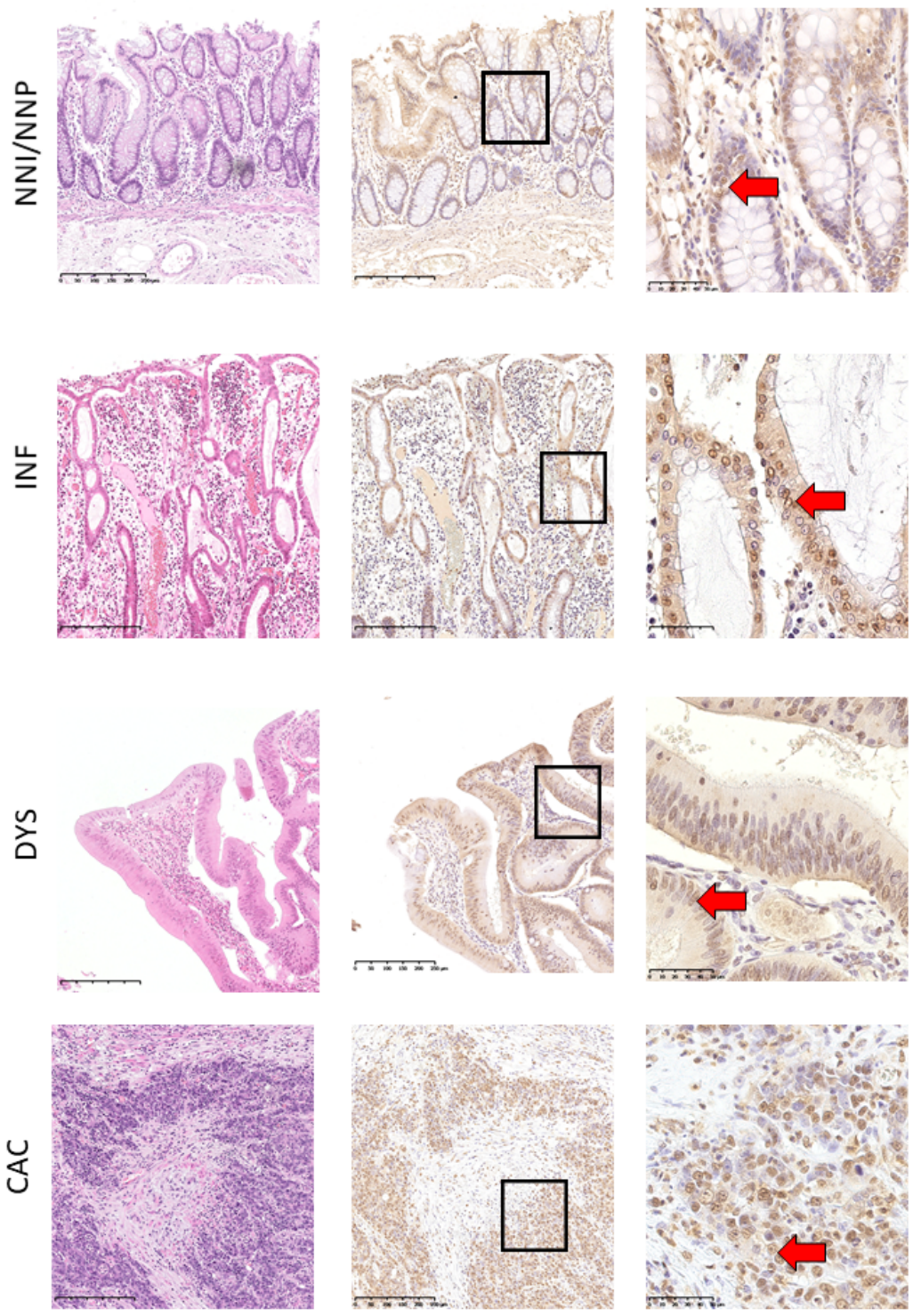



**Figure 35. Staining for pSMAD3 in the nuclei in different stages of UC and CAC**

Staining for pSMAD3 in different sections. A higher staining intensity is seen in the inflamed and caner tissues while a lower intensity is observed in the non-inflamed and the dysplastic epithelium. Red arrow indicates examples of stained nuclei of the epithelial cells. NNI/NNP- non-inflamed/ non-neoplastic, INF – inflamed, DYS – dysplastic, CAC – colitis associated cancer. 10X- scale bar - 250µm, 40X - scale bar - 50µm.

## 6.2 Expression of SMAD7 mRNA in different stages of CAC

### 6.2.1 Optimisation of *in-situ* hybridisation (ISH)

The RNAscope technique was first optimised using HeLa cells using a probe for isomerase B (PPIB) as positive control and a probe for the bacterial gene DapB as the negative control probe (**Figure 36**).

### 6.2.2 ISH for SMAD7 mRNA in different stages of CAC

The expression of SMAD7 mRNA was assessed using in-situ hybridisation with RNAScope. Sections that showed poor staining for the positive control (PPIB) due to time related degradation of mRNA had to be excluded. The expression of the mRNA normalised to the PPIB expression in 33 paraffin embedded sections was compared between different stages of CAC. When the ratio of SMAD7 mRNA molecules to PPIB was assessed there was no significant difference in the expression level in either stage compared to the non-inflamed epithelium (**Figures 37 to 39**).

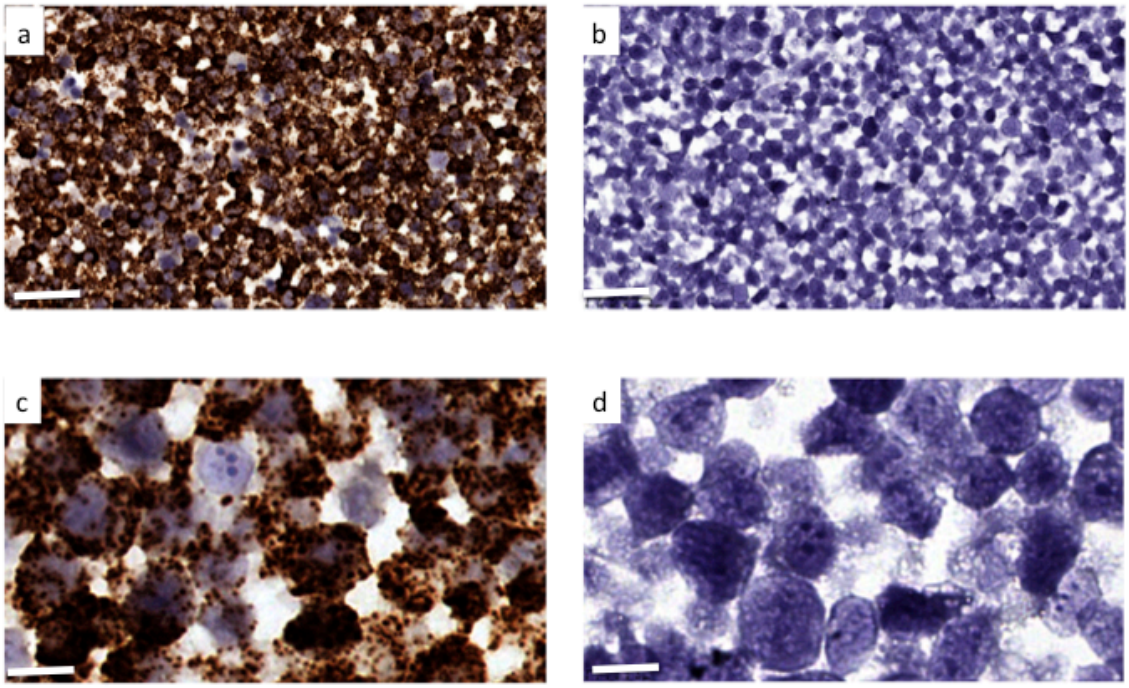



**Figure 36. Optimisation of ISH of SMAD7 mRNA**

RNAscope® technique was optimised on HeLa cells as per product recommendations a) 10X magnification with PPIB (positive probe) b) 10x magnification with DapB (negative probe). Scale bar - 250 µm. c) 40X magnification with PPIB and d) 40X magnification with DapB. Scale bar – 250 µm

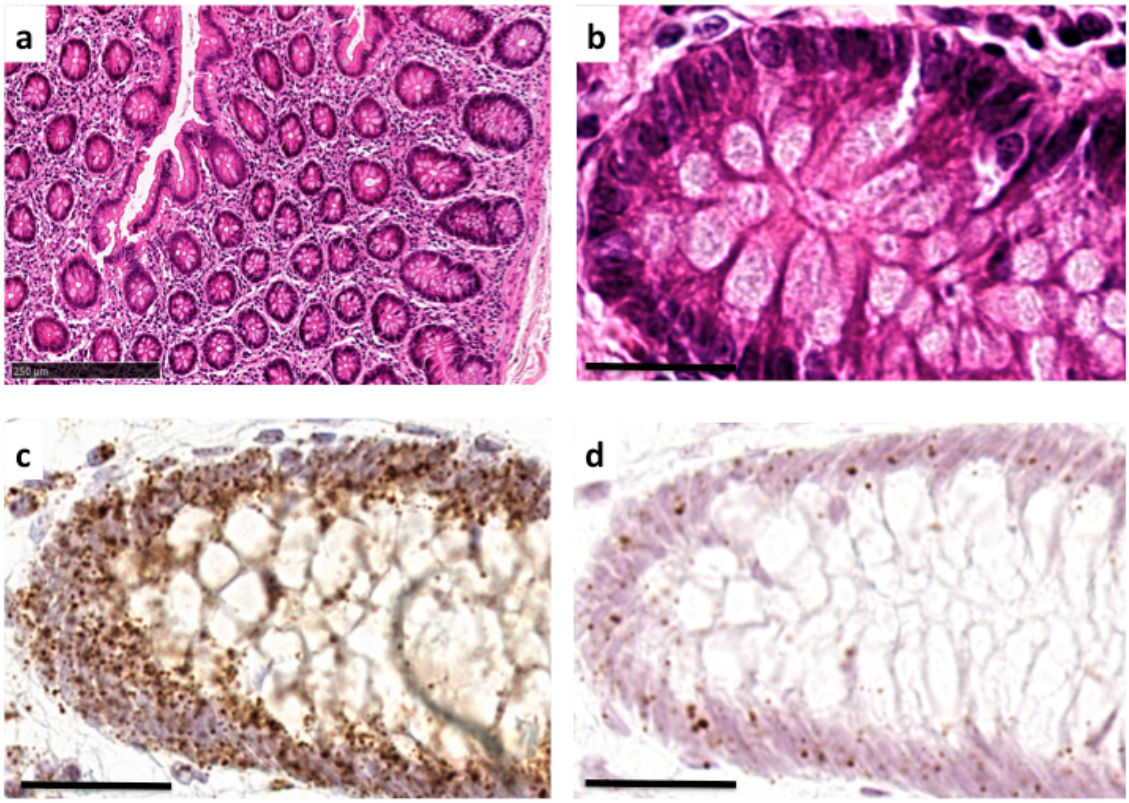



**Figure 37. In situ hybridisation of non-inflamed/ non-neoplastic colonic mucosa**

a) H&E staining of a colonic mucosa from a non-inflamed/ non neoplastic region, (10X) Scale bar - 250 µm, b) Magnified view (40X) of the H&E staining c) In-situ staining with the positive control demonstrating abundant mRNA molecules in the section. d) Serial section of the same site stained with SMAD7 mRNA. Scale bar - 250 µm.

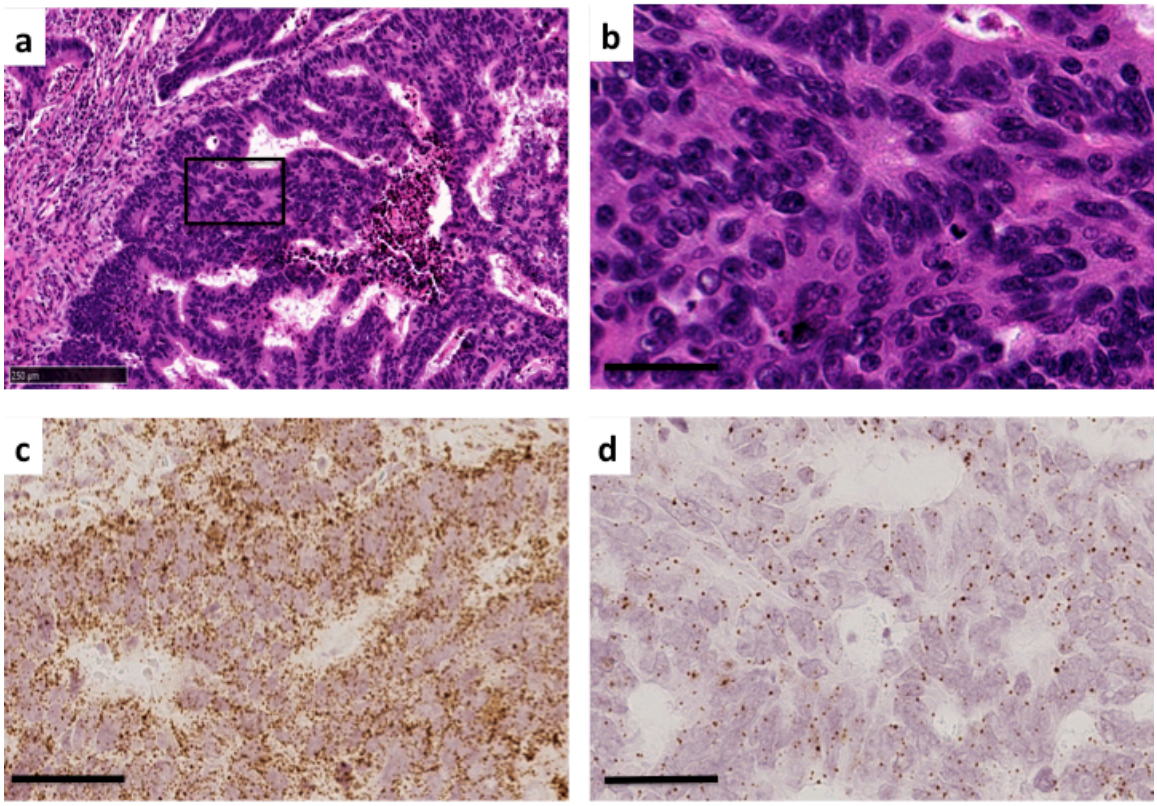



**Figure 38. In-situ hybridisation of malignant colonic mucosa**

a) H&E staining of a CAC (10X), Scale bar - 250 µm. b) Magnified view (40X) of the H&E staining, Scale bar - 50 µm c) In-situ staining with PPIB demonstrating abundant mRNA molecules in the section. 40X, Scale bar - 50 µm d) Serial section stained with SMAD7 ISH probe. Scale bar - 50 µm

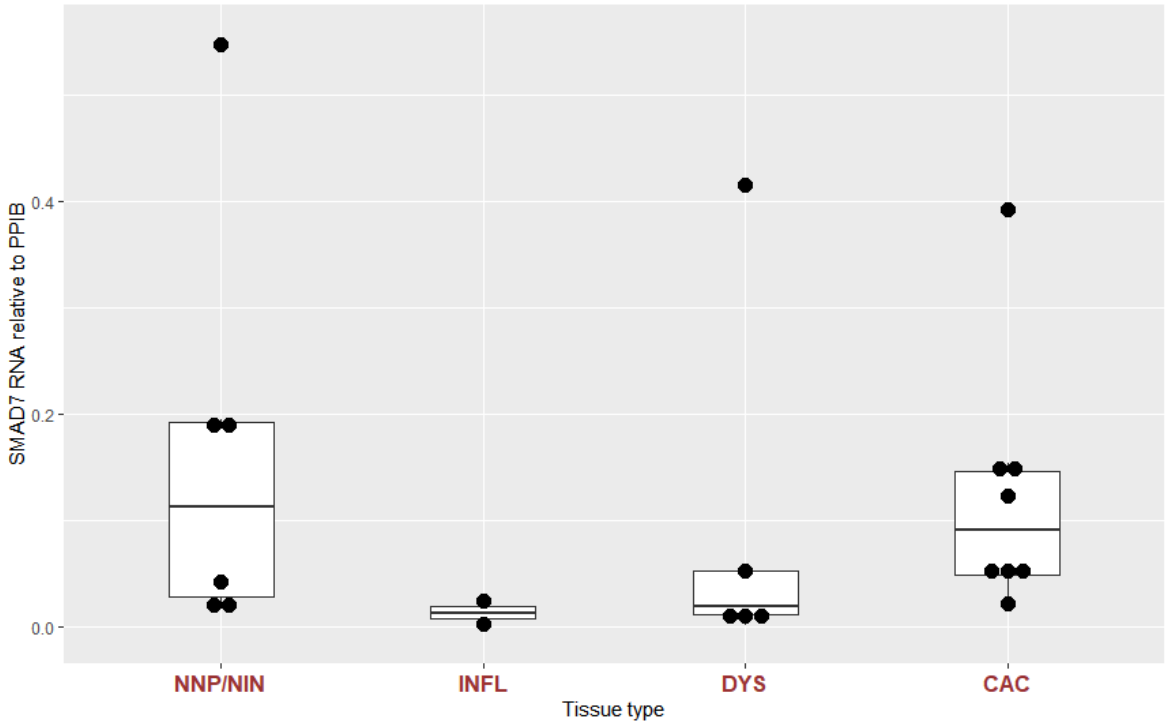



**Figure 39. In-situ hybridisation of SMAD7 mRNA**

Normalised expression levels of SMAD7 mRNA in non-neoplastic/ non-inflamed (NNP/NIN), inflamed (INF), dysplastic (DYS) and colitis associated cancer (CAC) tissue. The SMAD7 ISH was normalised in relation to PPIB (Peptidyl-prolyl cis-trans isomerase B). There was no significant difference in the ISH signal of SMAD7 mRNA in the epithelium of different tissue sections (p=0.34).

## 6.3 Methylation in SMAD7 genes influenced by SMAD in breast cancer

Analysing the WGBS data, I looked in to the promoter and genebodymethylation pattern of CDH1, CLDN4 and CGN; three geens that have been demnstrated to be differentially methylated influenced by SMAD7 in breast cancer (133). Promoter of CLDN4 demonstrated increasing levels of bypomethylation in dysplastic and cancer tissue although normal epithelium also showed hypomethylation of the promoter (**Figure 40**).

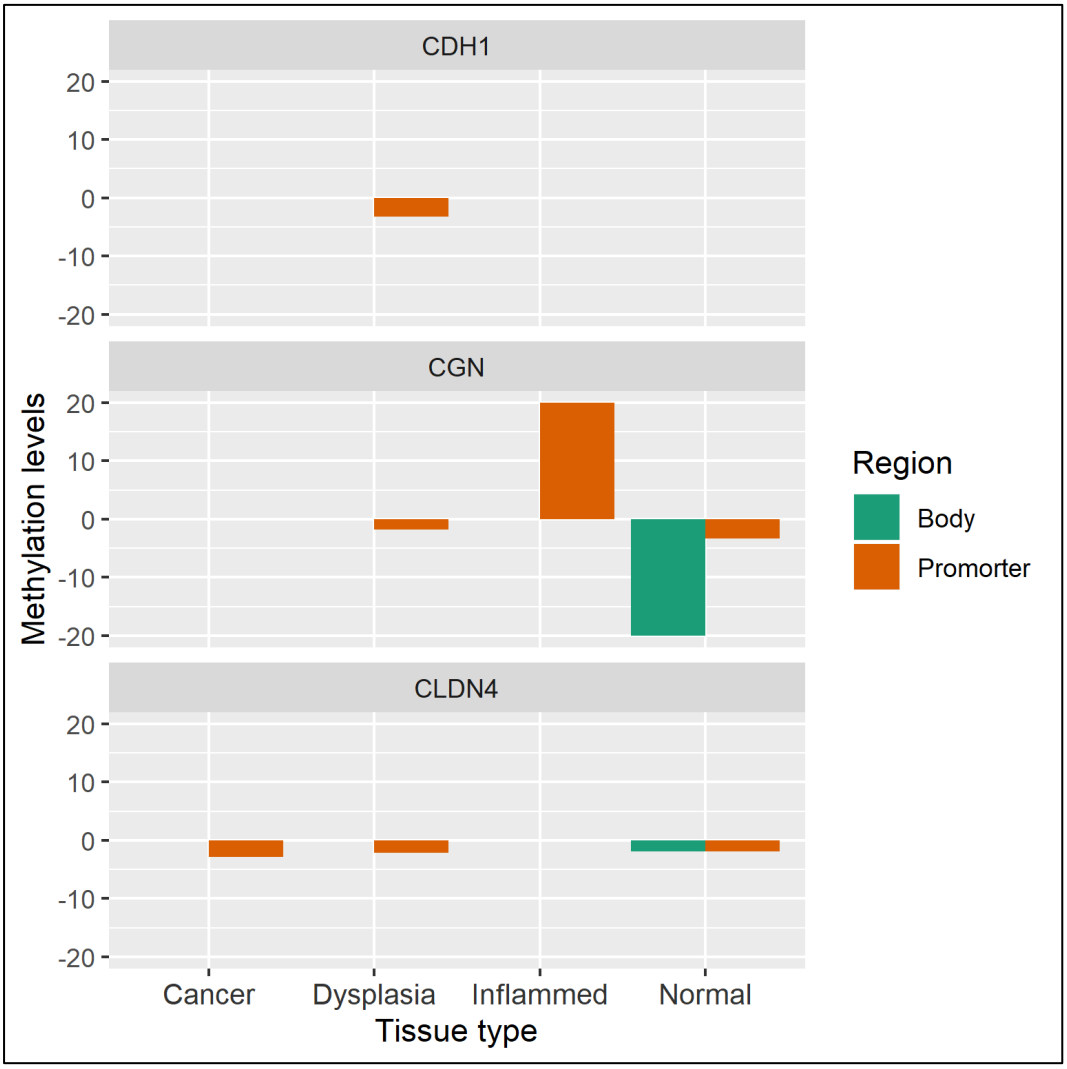



**Figure 40. Differential methylation of genes influenced by SMAD7**

Methylation pattern of CDH1, CGN and CLDN4 in different stages of CAC. There is no consistent pattern of methylation that can be distinguished from the normal epithelium or correlate to changes in SMAD7.

## 6.4 in-vitro knockdown of *SMAD7* gene in colorectal cancer cells

### 6.4.1 Gene knockdown with siRNA

No significant inhibition of *SMAD7* mRNA expression after treatment with the pooled siRNA was seen. RT-PCR failed to demonstrate any of the cell lines to have a difference of greater than 2 folds in gene expression (Figure 31).

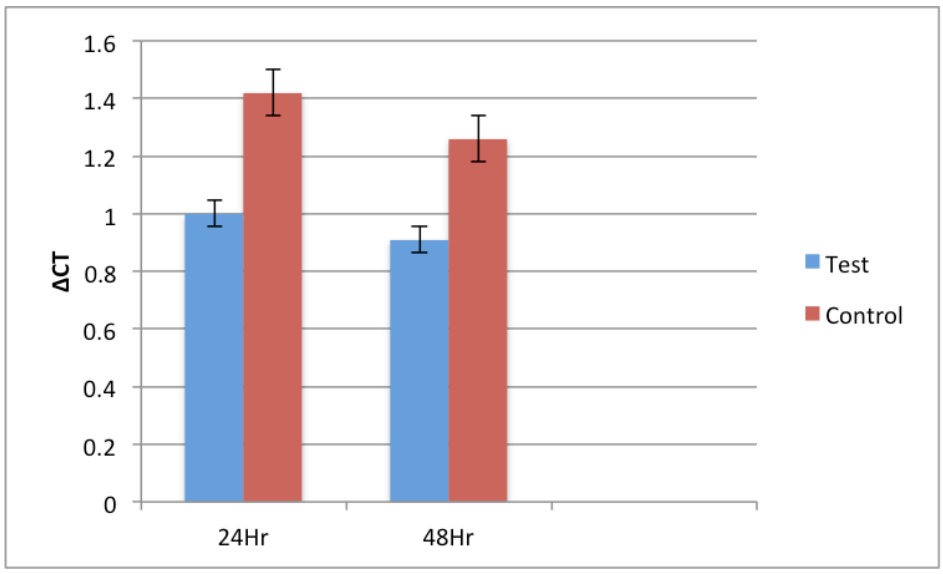



**Figure 31. *SMAD7* expression with siRNA transfection.**

RT-PCR at 24 and 48 hours after transfection of HCT116 cell lines with pooled siRNA. The transfected cells do not demonstrate a significant reduction in gene expression.

### 6.4.2 Gene knockdown with viral transfection of shRNA with Tet-ON construct

DLD-1 and HCT 116 colorectal cell lines were transfected with shRNA ligated to Tet-ON construct using a lentivirus vector. Two pairs of shRNA sequences compatible two sites of the SMAD7 gene in the non-variable region were custom made. The successful transfection was observed after positive puromycin selection. However,

both cell lines failed to demonstrate a significant reduction in the expression of *SMAD7* mRNA or SMAD7 protein (Figures 32 and 33).

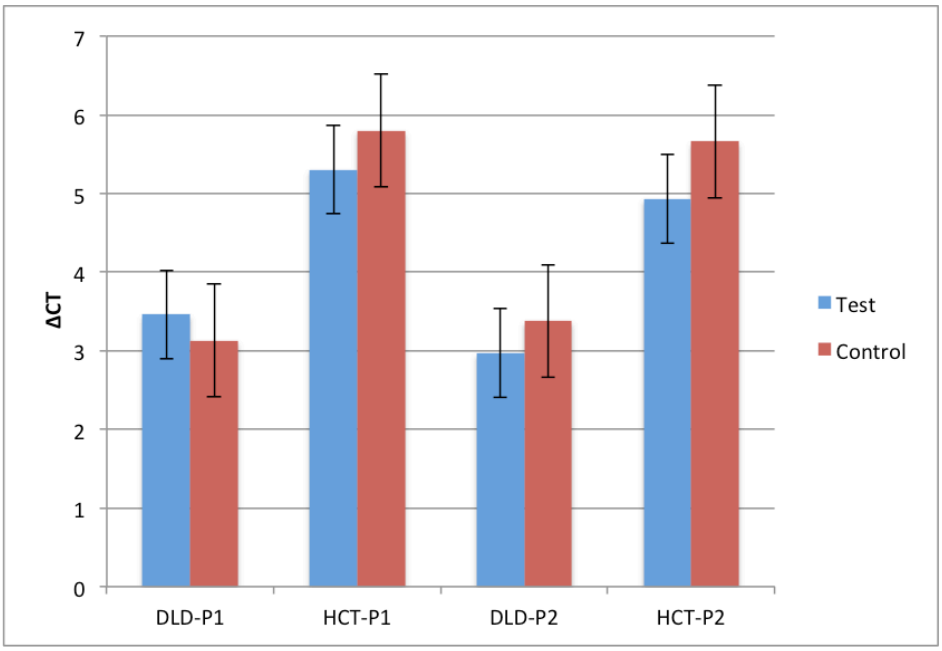



**Figure 32. SMAD7 expression with shRNA transfection.**

RT-PCR for *SMAD7* gene expression in HCT 116 and DLD-1 colorectal cancer cell lines transfected with *SMAD7* shRNA. None of the cell lines with neither of the shRNA sequences demonstrated a significant reduction in gene expression.

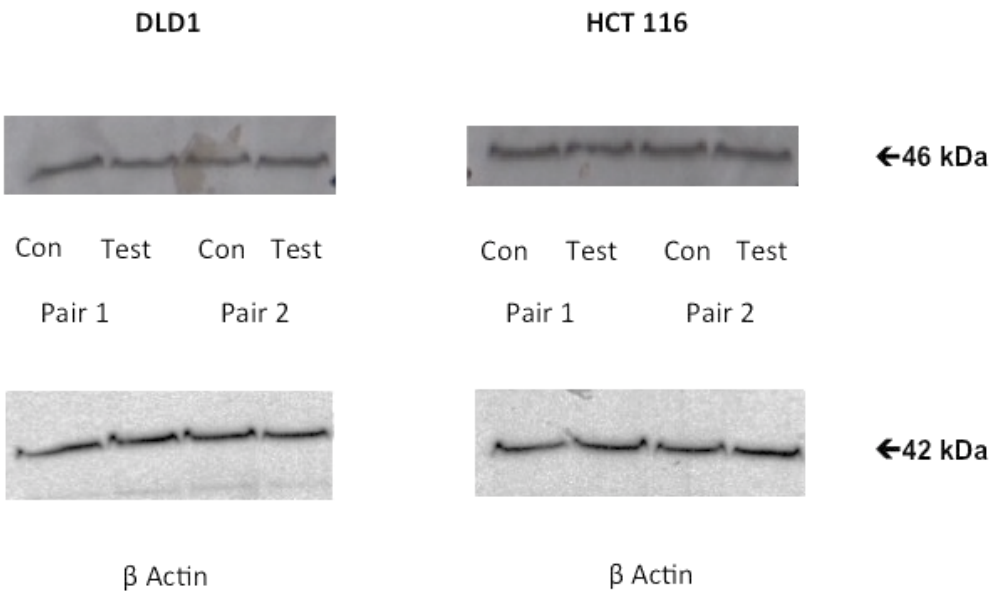



**Figure 33. SMAD7 protein expression after shRNA inhibition.**

Western blot detecting the expression of SMAD7 protein (46 kDa) in HCT 116 and DLD-1 cell lines following shRNA transfection. Expression of the SMAD7 protein in both cell lines was unaffected with the shRNA transfection.

## 6.5 Conclusion

There is a statistically significant bimodal increase in the expression of SMAD7 at different stages of UC associated cancer. As shown in Figures 12 to 15, inflamed and cancer tissue shows a significant increase in the expression of SMAD7 while non-inflamed and dysplastic epithelium shows a reduced expression of the same.

Although a complementary reduction in pSMAD3 expression is expected with an increase in SMAD7 expression, this was not to be observed. These results could be explained by the hypothesis of pSMAD3 escaping the inhibition of SMAD7 in the inflammatory and cancer epithelium. It has been suggested that the SMAD7 elevation is an initiating factor in UC. Given the dual action of TGFβ1 on cancer, the reduction in SMAD7 in the colonic epithelium during dysplastic phase may indicate

an attempt to allow the protective action of TGFβ1 in the precancerous stage. The increase in cancer tissue may represent the attempt to prevent the pro-carcinogenic action of TGFβ1. However, the absence of a complimentary reduction in pSMAD3 expression may indicate that TGFβ1 has a positive feedback response. This is the first time a bimodal pattern of SMAD7 expression has been described in the epithelium in UC and associated neoplasia.

It was also interesting to observe that SMAD7 mRNA levels did not differ significantly across different tissues. This could be due to the fact that increase in the SMAD7 protein levels in different tissue types is due to an altered metabolism of the protein through post transcriptional modifications rather than an increase in the transcription of the protein. Another possibility would be altered degradation of mRNA (279) increasing the total protein transcription while a constant amount of mRNA is maintained in the cytoplasm . This would require further studies to delineate the exact mechanism. When analysing the genes postulated to be influenced by SMAD7 (133) none of the genes demonstrated a correlation with the SMAD7 levels in different stages of CAC. Attempts at gene knockdown did not demonstrate a significant reduction in the SMAD7 mRNA levels in colorectal cancer cell lines. This could be either due to splice variances in the SMAD7 gene or low efficacy of the RNA constructs (280, 281). However successful transfer of the construct in to the cells was demonstrated with positive puromycin selection. The transcription of the SMAD7 mRNA may be more complex with several splice variances coding for similar peptides (280). These possibilities would need to be addressed in order to improve the results of this experiment. Addition of a reporter gene such as green fluorescent protein (GFP) could be used to monitor the transfection efficacy of siRNA and shRNA. Further use of novel technologies such as CRISPR could provide more

efficacious gene knockdown *in vitro*. A colonic organoid model with CRISPR technology to selectively inhibit SMAD7 would be a perfect experimental model to study the effect of SMAD7 on the methylation process.

# 7.0 Discussion

Ulcerative colitis, which is an autoimmune inflammation of the colonic mucosa due to, yet poorly explained aetiology, is increasing in incidence globally (6, 282). UC increases the risk of affected individuals to develop CAC. A dilemma faced by the clinicians is to predict which patients would progress to develop CAC. The main aim of this study was the characterisation of the epigenome of CAC, which would allow the identification of potential biomarker for the malignancy. As the process of CAC development occurs in a staged manner from inflammation to dysplasia further transforming into cancer, a potential methylation marker that could predict the sequel early would significantly change the management of the disease and aid the decision to offer early preventive surgery. SMAD7 is a molecule that has been found to be elevated in inflamed colonic epithelium and has been tried as a therapeutic target. SMAD7 has been implicated in the development of UC and CRC cancer by many researchers. However the changes of SMAD7 were not studied in CAC.

In the first part of my research, whole genome bisulphite sequencing was performed for tissues from different stages of CAC. Previously array based methylation studies have been performed for CAC samples and inflamed tissues but a comparison across all stages of CAC along with normal colonic mucosa has not been performed in a single study. Coupled controls were included for each test sample from the same patient. This was to exclude the age-related and inherited methylation changes and to identify the differentially methylated genes that are unique to each stage. Furthermore, since adjacent tissue was taken as control samples for dysplasia and CAC, the differentially methylated genes related to inflammation alone are excluded from the analysis. Since the objective of the project was to provide guidance for potential methylation markers to predict the development of CAC, common genes showing similar methylation pattern across all stages were searched for.

Overall CAC tissue demonstrated a higher proportion of hypomethylated promoters compared to other stages. There were 63 common genes that were hypomethylated in inflamed, dysplastic and cancer epithelium compared to their matched control which were not observed in normal epithelium. Out of these, 7 candidate genes were selected through available evidence to link them with CRC. These 7 genes; TACSD2, eIF6, PRDX1, USP1, IRAK-3, LFNG and S100A12 were strongly linked to CRC and were unlikely to be associated inflammatory disease alone. described by previous authors .In particular, TACSD2 (Tumor-associated calcium signal transducer 2) is known to be upregulated in CRC and increased metastatic potential is linked in to it (199, 283-286). The gene codes for a calcium signal transducer that has shown to be over expressed in epithelial cancers including colon cancer (287). The gene product, Trop-2 is demonstrated to be important in cell adhesion and maintaining the epithelial barrier . Trop-2 has been proposed as a biomarker and a molecular target in cancers including CRC (287). Fact that TACSD2 promoter is hypomethylated in the UC affected tissues from the early stages, in contrast to healthy epithelium may imply its potential role in the development of CAC and its use as a early biomarker.

Product of eIF6 is a protein implicated in translational initiation and in ribosome synthesis, is linked to CRC proliferation and invasion (200-202). This protein is highly expressed in CRC compared to normal tissue. It is involved in the positive regulation of AKT-related cancer signalling and it enhances tumorigenicity in CRC. It is also known to activate aberrant Wnt/β-catenin signalling pathways in cancer tissues (202). In this study all three stages of CAC showed promoter hypomethylation of eIF6 compared to matched control samples, which can be related to its overexpression, and the same was not observed in normal epithelium. Interestingly it

is being observed in sporadic CRC that eIF6 is increasingly expressed in dysplastic adenomas and cancers in contrast to normal tissue (202, 288). In this study it is observed that promoter hypomethylation is present form the inflammatory stage of UC that indicates its strong potential as an early biomarker.

Relationship between chronic inflammation and CRC has long been identified (203). PRDX1 gene, which codes for a thiol-peroxidase protein that degrades hydroperoxides to water, is known to be upregulated in CRC tissue compared to normal tissue (204, 205). The inactivation of PRDX1 inactivation has a positive effect on activating pro-carcinogenic pathways such as nuclear factor-κB (NF-κB) (203-205). Hence hypermethylation of PRDX1 from the stage of inflammation onwards may suggest that this could be used as an early marker of neoplastic change.

USP1, a recently identified gene of interest in CRC, is also amongst the hypomethylated common genes across all stages and its expression was validated with qPCR data. USP1 is over expressed in CRC and its inhibition slows down HCT116 cell growth (206-208). Promoter hypomethylation of such a gene that is strongly associated with CRC shows high potential to be a marker of malignant transformation in UC.

IRAK3 gene which is shown to be upregulated in CRC (289) has been observed to be down regulated in murine models of CAC (209). Although these contrasting findings have not been validated we observed it to be uniformly hypomethylated along all stages of CAC in this study highlighting its potential as a marker. It is postulated that IRAK3 is an inhibitor of inflammation and has a pro-carcinogenic effect on the colonic epithelium. Rothschild et al and other investigators have repeatedly demonstrated that IRAK3 in humans favours the development and progression of colon cancer (289, 290) . It is shown to act via combined Wnt and

TLR activation in the colon cancer cells with associated poor prognosis in patients (289). A gene that produces a molecule, which activates the Wnt pathway showing promoter hypomethylation is a significant finding and may indicate the early inclination of inflamed mucosa in UC towards tumourogenecity.

The expression of LFNG, which codes for Lunatic fringe a receptor ligand of the N-acetyl-glucosaminyl-transferases family, is upregulated in CRCs and showed promoter hypomethylation in all stages of CAC in the current study (211-213). Interestingly LFNG is known to be upregulated in a TGFβ dependent manner in CRC cell lines (211). LFNG is known to both positively and negatively modulate Notch signalling, which is known to be active in CRC. Promoter hypomethylation of LFNG was observed in all stages of CAC and its upregulated expression was confirmed with qPCR data in this study. These data proposes a strong case for this gene to be included in a candidate gene panel for an early biomarker although the qPCR in this study could not confirm increased expression levels of the gene.

The increased expression of *S100A12* has been observed in a number of chronic inflammatory conditions including IBD (291-293). Interestingly the same has been detected to be highly expressed in CRC and has been suggested as a faecal biomarker for the early detection of CRC (214, 294). A high specificity of 95% was observed for the marker pair S100A12 and haemoglobin-haptoglobin (214). Promoter hypomethylation of *S100A12* gene was identified in all stages of CAC in the study samples. Being related to both chronic inflammation and CRC, this marker may be a strong link between inflammation and cancer in CAC.

Four hypermethylated were identified as linked to colorectal cancer from the analysis of differentially methylated genes. These four were consistent across all stages and was not observed in the normal colonic epithelium. Amongst the

hypermethylated genes *ICAM5*, an adhesion molecule, and *IGF2*, a growth factor receptor, are both linked to CRC (218, 221, 222). *ICAM5* hypermethylation leads to reduction in cell-cell adhesion increasing the invasive and metastatic potential (217). Both up-regulation and inhibition of IGF2 has been observed in CRC and its specific action has not yet been clearly understood (222).

Another gene with hypermethylated promoter, *MIR-124*, is a gene previously shown to be down regulated in CRC. It is a tumour suppressor gene with multiple actions including targeting STAT3 pathway to inhibit inflammation, changing genetic propensity, inhibiting Warburg effect, inducing natural cell death and autolysis of cells (224, 295, 296). It is shown to be methylated during the carcinogenic process in sporadic CRC (295, 297-299). MIR-124 is also suggested as a therapeutic target in CRC by increasing the radio sensitivity of rectal cancers (296). An interesting link between inflammation and colorectal cancer is demonstrated through the inhibitory action of MIR-124 on the pro-inflammatory molecule STAT3, which is over expressed in UC (300). Hypermethylation of MIR-124, a gene with multiple tumour suppressor activities, in all stages of CAC is a significant indicator that the pro-carcinogenic factors may be active from the inflammatory stage in UC. Therefore methylation of MIR-124 provides a strong case as an early biomarker. .

Promoter hypermethylation of WT1, a tumour suppressor gene first identified in Wilms' tumours, is a significant finding. Promoter methylation of WT1 has repeatedly been demonstrated in sporadic CRC and other epithelial cancers (249, 301). Magnani et al studying young (< 40 years) and older populations (>60 years) with CRC found promoter methylation of WT1 to be a consistent finding in both populations (302). A tumour suppressor that acts through inhibiting the Wnt pathway

to be methylated from the stage of inflammation throughout all stages of CAC increases its potential as an early biomarker of colon cancer in UC.

I then analysed the pattern of gene body methylation across all stages of CAC. Gene body methylation although shown to correlate with the expression levels of the gene, this correlation is variable (227). Out of the 465 common genes with body methylation, 92 genes with the highest level of hypomethylation and hyper methylation in cancer, or both cancer and dysplasia were identified. Out of the 10 hypomethylated genes with direct links to CRC, five (*ADAM23, ITGA8, ZEB2, ETS1* and *CDH12*) were either cell adhesion molecules or integrins involved in epithelial mesenchymal transition (Table 12). Others were transcription factors well described in CRC. The two genes related to CRC carcinogenesis, showing gene body hypermethylation are *WT1-AS*, a gene coding for an anti-sense RNA up stream of *WT1*, and *CABIN1* gene, which is involved in apoptosis. These genes were all differentially methylated from the inflammatory stage to cancer compared to the control samples. Few of these genes require special mention due to their strong relationship with CRC. *ETS1* and *ZEB2* are both transcription initiation factors and both genes are linked to invasiveness of CRC cells (235-237, 242, 303). *ETS1*, which is categorised a proto-oncogene, has shown to control the expression of genes involved in extracellular matrix remodelling and is highly expressed in CRC and adenomas with high-grade dysplasia (242). Similarly *ZEB2*, which regulates epithelial mesenchymal transition, has been proposed as a molecular marker for tumour stage classification in post-surgical specimens of CRC (236, 237). Both these genes are hypomethylated in the study samples making them strong contenders to be biomarkers with the validation of their expression levels with qPCR. Products of *HPP1* gene, which act as transmembrane receptors, has growth factor like effects

and is shown to be hypermethylated in CAC by Sato et al (304). Genes with such pivotal roles in the development of CRC being differentially methylated from the inflammatory stage may suggest their predictive value in the development of CAC. *USP1*, is interestingly also amongst the genes with highest level of gene body hypomethylation. Having both promoter and gene body hypomethylation in this gene in all stages of CAC is of significance. However the gene expression data could not confirm an increased level of expression of the gene with qPCR.

Both ITGA8 and CDH12 are genes producing in integrins and cell adhesion molecules and have been strongly indicated in tumorogenicity of CRC (228, 229, 233, 234). Both these genes show increasing levels of gene body hypomethylation along the stages of CAC with maximum level of hypomethylation in cancer.

In the next part of my research I looked into the methylation of key driver genes identified in sporadic CRC. Analysing methylation level of promoter and gene body of *APC, KRAS, BRAF, PIK3CA, SMAD4* and *p53* it was observed that none of them showed significant methylation pattern at different stages of CAC. *BRAF* demonstrated hypomethylation of both the gene body and the promoter while the *P53* showed hypomethylation of the gene body. BRAF is an oncogene that has been found to be activated in around 30% of CRC and P53 is a tumour suppressor gene usually inactivated through mutation in around 50% of the CRC.

Further genes that have been suggested to be methylated in CAC previously were analysed. Eleven genes; *RUNX3, MINT1, MYOD1, APC2, P16 (CDKN2A), SFRP2, SFRP4, SFRP5, EYA4, ESR* and *WIF1* were analysed to look for promoter or gene body methylation. P16 (CDKN2A), SPRF2 and MYOD1 were the only genes, which demonstrated differentially methylated promoters in cancer or dysplasia that were not present in the normal tissue. Promoter hypermethylation of P16 (CDKN2A) was

observed in inflammatory and malignant tissue in contrast to hypomethylation in normal tissue. It has been shown to be methylated in the promoter region in dysplastic and cancer tissue associated with UC (268, 305). Its product, a cyclin - dependent-kinase, is a cell cycle regulatory molecule, absence of which inhibits apoptosis (272, 306). The dysplastic samples of my study were not characterised by significant differential methylation of P16 compared to adjacent tissue. This may be comparable to previous observations where it was demonstrated that P16 methylation was observed in both neoplastic and non-neoplastic tissue in UC (268). Increased expression of P16 has also been has been reported at single crypt level in CAC (273). *MYOD1* is known show age related methylation in colonic mucosa in normal individuals (307) . SFRP2 another gene involved in the Wnt signalling pathway has shown to be methylated in CRC with a potential to be used as a biomarker (308). These three genes although not demonstrating a uniform pattern across the stages of CAC, they have a potential of being used as predictors with the available evidence. SEPTIN9 codes for proteins involved in cell division and has been identified as a biomarker in CRC (309). It has been detected to be methylated in CRC and has been isolated from both blood and stools as a sensitive biomarker (310, 311). In this study SPETIN9 was found to be hypomethylated in inflamed and dysplastic tissue in contrast to normal tissue. However a differential methylation was not demonstrated in CAC tissue compared to adjacent non-neoplastic tissue. Given its distinction in premalignant UC affected tissue and promising data as a methylation marker in sporadic CRC, SPETIN9 was included in the proposed candidate gene set. The GO analysis of the selected markers further supports their potential to be effective predictors of CAC. The gene set is annotated to three significant molecular functions; cell adhesion molecule binding and integrin binding. Further, the 15

hypomethylated genes from the candidate set are annotated to cell adhesion molecule binding, cadherin binding and MAP kinase activity which are all closely associated to the process of carcinogenesis. A set of gene annotated to such processes being differentially methylated from the stage of inflammation and being absent in normal colonic tissue is a significant phenomenon. Furthermore, the gene set is annotated to a number of disease processes, which include UC, colonic polyps, epithelial cancers and sarcomas. This further justifies the potential of these genes to be carried forward to the next level of validation.

This is the first attempt at looking at differentially methylated genes common to all stages of CAC using WGBS data from isolated epithelial cells from colonic samples. Furthermore the samples are from different patients and surrounding non-neoplastic tissues were used as control samples making them more specific markers.

In the second part of my research I have demonstrated how the SMAD7 molecule expression changes along the different stages of CAC. The bimodal change in expression of SMAD7 was not previously demonstrated. Several hypotheses can be drawn out of the observation as the reason behind this phenomenon, which needs to be tested. Lower SMAD7 expression in the dysplastic stage could be as attempt to increase the anti tumour effect of TGFβ1 during early stage of CAC. The second peak in the cancer tissue could be an attempt to inhibit TGFβ1 action and its pro-carcinogenic effect in late stages as explained previously(138, 147, 149, 312). Previously an increase of SMAD7 in the immune cells has been observed (125, 144), but there were no studies focusing on the change in SMAD7 in colonic epithelial cells. It was also noted that the expression of SMAD7 mRNA does not reflect the changes seen in protein within the epithelial cells. This might be due to a regulation in metabolism as explained earlier (141). However no methylation changes were

detected in the SMAD7 gene or a significant differential methylation could be detected in the known genes that are affected by SMAD7. The attempt at silencing the SMAD7 mRNA in colorectal cancer cell lines did not demonstrate a significant reduction at protein levels despite the successful transfer of pLKO-Tet-ON constructs in to the cell lines, which was confirmed by puromycin selection. A possible explanation for this could be due to a splice variance in SMAD7 mRNA that allowed it to by pass the inhibition of shRNA. An ideal model to overcome this would be using CRISPR technology on a colonic organoid model. There were both time and resource constrains to organise such an experiment during this project.

In conclusion, in this pilot study I have identified 27 genes that could be used in a potential gene panel to predict CAC development in UC. Out of these, 11 genes show differentially methylated promoters while 12 genes show gene body DMRs. Twenty-three of the genes consistently show differential methylation in all three stages of CAC development in contrast to the normal colonic epithelium. The other 4 genes, P16 (CDKN2A), SFRP2, MYOD1 and SEPTIN9 were selected from already identified CAC and CRC biomarkers with potential. The selected genes are annotated to important molecular functions related to cancer as well as UC, colonic polyps and other epithelial cancers in disease process annotation.

A limitation of my study is the limited number of samples used in the sequencing. CAC samples are not abundant in clinical practice and to overcome this I have used a combination of both fresh and FFPE tissue. However utilising WGBS and using adjacent tissue as control sample I attempted to increase the specificity of my data set. In addition laser capturing epithelial cells helped to increase the specificity of the data by excluding the mesenchymal tissue in the samples. Difficulty in acquiring

adequate amount of DNA from isolated epithelial cells with this method also limited the number of samples I could utilise during the study period.

Adjacent tissue and buffy coat were the best available tissue for control samples. Since the objective was to detect the acquired CpG island methylation, changes due to the inflammation age-related and germline mutations had to be excluded from the test samples. However, there may be variations in methylation between these test and control samples that are independent of the disease process. Epithelial cells of the sigmoid colon taken as a control sample against a test sample of a rectal cancer may harbour site-specific differences in DNA methylation. Several authors have demonstrated differences in methylation in normal colon tissue from left and right colons (313-315). Barnicle and colleagues demonstrated differences in methylated regions between proximal and distal colon tissues. Also, interesting to note was the increased variability in methylation patterns in whole biopsies compared to isolated epithelial cells in their study (314). Using isolated epithelial cells in the current study would reduce the intrasite variations in methylation levels. The best control sample would be serial samples collected at different time points from the same site. This would only be possible in a long-term prospective study where subjects will be followed up with serial screening colonoscopies with biopsies. The inflamed tissues in this study were matched with buffy coat as the entire colon was inflamed and a suitable adjacent area was not available as a control. The buffy coat would allow to subtract environmentally influenced and inherited methylation from the test samples, but there are site specific methylation changes that would not be compared (257). However, notably cancer driving epigenetic changes have shown not to correlate between primary tumour sites and buffy coat (256).

I envisage validating the selected genes in a prospectively collected larger cohort of CAC related tissue samples in order to assess their predictive value in a clinical setting. Fresh tissue collected from patients with inflammed, dysplastic and malignant colonic lesions during colonoscopy could be used to assess the methylation levels in mucosal samples. Also same samples could be used to validate the expression levels of the identified gene products using qPCR. Longitudinal follow-up data could be correlated with the findings of the methylation studies and differentiate the progressors to CAC from non-progressors. Also studying the methylation of the proposed genes in a larger cohort of patients could help on narrowing down the number of candidate genes. Further in-vitro studies to delineate the exact mechanism of some of the methylation would be of help. An organoid model for UC, if developed, could be used to identify the exact effect of methylation of the proposed genes. Further such a model could be used to study the action of SMAD7 on methylation and development of CAC by utilising techniques such as CRISPR to selectively inactivate the gene.

Development of a validated gene panel with these genes could assist in early decision making for colectomy in patients with UC minimising the cost, risk of complications and development of cancer.

# 8.0 References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017;66(4):683-91.

2. Center MM, Jemal A, Smith RA, Ward E. Worldwide variations in colorectal cancer. CA Cancer J Clin. 2009;59(6):366-78.

3. De Rosa M, Pace U, Rega D, Costabile V, Duraturo F, Izzo P, et al. Genetics, diagnosis and management of colorectal cancer (Review). Oncol Rep. 2015;34(3):1087-96.

4. Sebastian S, Hernandez V, Myrelid P, Kariv R, Tsianos E, Toruner M, et al. Colorectal cancer in inflammatory bowel disease: results of the 3rd ECCO pathogenesis scientific workshop (I). J Crohns Colitis. 2014;8(1):5-18.

5. Burisch J, Jess T, Martinato M, Lakatos PL, EpiCom E. The burden of inflammatory bowel disease in Europe. Journal of Crohn's & colitis. 2013;7(4):322-37.

6. Ng SC, Tang W, Ching JY, Wong M, Chow CM, Hui AJ, et al. Incidence and Phenotype of Inflammatory Bowel Disease Based on Results From the Asia-Pacific Crohn's and Colitis Epidemiology Study. Gastroenterology. 2013;145(1):158-65.e2.

7. Niriella MA, De Silva AP, Dayaratne AHGK, Ariyasinghe MHADP, Navarathne MMN, Peiris RSK, et al. Prevalence of inflammatory bowel disease in two districts of Sri Lanka: a hospital based survey. BMC gastroenterology. 2010;10:32-.

8. Collaborators GBDIBD. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol. 2020;5(1):17-30.

9. The American Society of Colon and Rectal Surgeon's Annual Meeting Abstracts. Diseases of the Colon & Rectum. 2018;61(5).

10. Abegunde AT, Muhammad BH, Bhatti O, Ali T. Environmental risk factors for inflammatory bowel diseases: Evidence based literature review. World journal of gastroenterology. 2016;22(27):6296-317.

11. Ananthakrishnan AN. Epidemiology and risk factors for IBD. Nature Reviews Gastroenterology & Hepatology. 2015;12(4):205-17.

12. Loddo I, Romano C. Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. Frontiers in immunology. 2015;6:551-.

13. Low D, Nguyen DD, Mizoguchi E. Animal models of ulcerative colitis and their application in drug research. Drug Des Devel Ther. 2013;7:1341-57.

14. Lennon EM, Maharshak N, Elloumi H, Borst L, Plevy SE, Moeser AJ. Early life stress triggers persistent colonic barrier dysfunction and exacerbates colitis in adult IL-10-/- mice. Inflamm Bowel Dis. 2013;19(4):712-9.

15. Smith F, Clark JE, Overman BL, Tozel CC, Huang JH, Rivier JE, et al. Early weaning stress impairs development of mucosal barrier function in the porcine intestine. Am J Physiol Gastrointest Liver Physiol. 2010;298(3):G352-63.

16. Kang M, Martin A. Microbiome and colorectal cancer: Unraveling host-microbiota interactions in colitis-associated colorectal cancer development. Semin Immunol. 2017;32:3-13.

17. Ordas I, Eckmann L, Talamini M, Baumgart DC, Sandborn WJ. Ulcerative colitis. Lancet. 2012;380(9853):1606-19.

18. Van Klinken BJ, Van der Wal JW, Einerhand AW, Buller HA, Dekker J. Sulphation and secretion of the predominant secretory human colonic mucin MUC2 in ulcerative colitis. Gut. 1999;44(3):387-93.

19. Heller F, Florian P, Bojarski C, Richter J, Christ M, Hillenbrand B, et al. Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. Gastroenterology. 2005;129(2):550-64.

20. Heller F, Fromm A, Gitter AH, Mankertz J, Schulzke JD. Epithelial apoptosis is a prominent feature of the epithelial barrier disturbance in intestinal inflammation: effect of pro-inflammatory interleukin-13 on epithelial cell function. Mucosal Immunol. 2008;1 Suppl 1:S58-61.

21. Consortium UIG, Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet. 2009;41(12):1330-4.

22. Wheeler JM, Kim HC, Efstathiou JA, Ilyas M, Mortensen NJ, Bodmer WF. Hypermethylation of the promoter region of the E-cadherin gene (CDH1) in sporadic and ulcerative colitis associated colorectal cancer. Gut. 2001;48(3):367-71.

23. Yaeger R, Shah MA, Miller VA, Kelsen JR, Wang K, Heins ZJ, et al. Genomic Alterations Observed in Colitis-Associated Cancers Are Distinct From Those Found in Sporadic Colorectal Cancers and Vary by Type of Inflammatory Bowel Disease. Gastroenterology. 2016;151(2):278-87.e6.

24. Parker B, Buchanan J, Wordsworth S, Keshav S, George B, East JE. Surgery versus surveillance in ulcerative colitis patients with endoscopically invisible low-grade dysplasia: a cost-effectiveness analysis. Gastrointest Endosc. 2017;86(6):1088-99 e5.

25. Bach SP, Mortensen NJ. Ileal pouch surgery for ulcerative colitis. World J Gastroenterol. 2007;13(24):3288-300.

26. Fazio VW, Kiran RP, Remzi FH, Coffey JC, Heneghan HM, Kirat HT, et al. Ileal Pouch Anal Anastomosis. Annals of Surgery. 2013;257(4):679-85.

27. Kirat HT, Remzi FH. Technical aspects of ileoanal pouch surgery in patients with ulcerative colitis. Clin Colon Rectal Surg. 2010;23(4):239-47.

28. McIntyre BP, Pemberton JH, Wolff BG, Beart RW, Dozois RR. Comparing functional results one year and ten years after ileal pouch-anal anastomosis for chronic ulcerative colitis. Diseases of the Colon & Rectum. 1994;37(4):303-7.

29. Venkatesh PG, Jegadeesan R, Gutierrez NG, Sanaka MR, Navaneethan U. Natural history of low grade dysplasia in patients with primary sclerosing cholangitis and ulcerative colitis. J Crohns Colitis. 2013;7(12):968-73.

30. Choi C-HR, Rutter MD, Askari A, Lee GH, Warusavitarne J, Moorghen M, et al. Forty-Year Analysis of Colonoscopic Surveillance Program for Neoplasia in Ulcerative Colitis: An Updated Overview. The American Journal of Gastroenterology. 2015;110(7):1022-34.

31. Rutter M, Saunders B, Wilkinson K, Rumbles S, Schofield G, Kamm M, et al. Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. Gastroenterology. 2004;126(2):451-9.

32. Collins P, Rhodes J. Ulcerative colitis: diagnosis and management. BMJ. 2006;333(7563):340-3.

33. Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. Gut. 2001;48(4):526-35.

34. Jess T, Horváth-Puhó E, Fallingborg J, Rasmussen HH, Jacobsen BA. Cancer Risk in Inflammatory Bowel Disease According to Patient Phenotype and Treatment: A Danish Population-Based Cohort Study. The American Journal of Gastroenterology. 2013;108(12):1869-76.

35. Kinugasa T, Akagi Y. Status of colitis-associated cancer in ulcerative colitis. World journal of gastrointestinal oncology. 2016;8(4):351-7.

36. Moss SF, Blaser MJ. Mechanisms of disease: Inflammation and the origins of cancer. Nature clinical practice Oncology. 2005;2(2):90-7; quiz 1 p following 113.

37. Okayasu I. Development of ulcerative colitis and its associated colorectal neoplasia as a model of the organ-specific chronic inflammation-carcinoma sequence. Pathology International. 2012;62(6):368-80.

38. Dyson JK, Rutter MD. Colorectal cancer in inflammatory bowel disease: what is the real magnitude of the risk? World J Gastroenterol. 2012;18(29):3839-48.

39. Ekbom A, Helmick C, Zack M, Adami HO. Ulcerative colitis and colorectal cancer. A population-based study. N Engl J Med. 1990;323(18):1228-33.

40. Jess T, Rungoe C, Peyrin-Biroulet L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. Clin Gastroenterol Hepatol. 2012;10(6):639-45.

41. Jess T, Simonsen J, Jorgensen KT, Pedersen BV, Nielsen NM, Frisch M. Decreasing risk of colorectal cancer in patients with inflammatory bowel disease over 30 years. Gastroenterology. 2012;143(2):375-81 e1; quiz e13-4.

42. Damas OM, Abreu MT. Are patients with ulcerative colitis still at increased risk of colon cancer? Lancet. 2020;395(10218):92-4.

43. Aust DE, Terdiman JP, Willenbucher RF, Chang CG, Molinaro-Clark A, Baretton GB, et al. The APC/beta-catenin pathway in ulcerative colitis-related colorectal carcinomas: a mutational analysis. Cancer. 2002;94(5):1421-7.

44. Johnson DH, Taylor WR, Aboelsoud MM, Foote PH, Yab TC, Cao X, et al. DNA Methylation and Mutation of Small Colonic Neoplasms in Ulcerative Colitis and Crohn's Colitis. Inflammatory Bowel Diseases. 2016;22(7):1559-67.

45. Arcila M, Lau C, Nafa K, Ladanyi M. Detection of KRAS and BRAF mutations in colorectal carcinoma roles for high-sensitivity locked nucleic acid-PCR sequencing and broad-spectrum mass spectrometry genotyping. The Journal of molecular diagnostics : JMD. 2011;13(1):64-73.

46. Leedham SJ, Graham TA, Oukrif D, McDonald SA, Rodriguez-Justo M, Harrison RF, et al. Clonality, founder mutations, and field cancerization in human ulcerative colitis-associated neoplasia. Gastroenterology. 2009;136(2):542-50 e6.

47. Schulmann K, Mori Y, Croog V, Yin J, Olaru A, Sterian A, et al. Molecular phenotype of inflammatory bowel disease-associated neoplasms with microsatellite instability. Gastroenterology. 2005;129(1):74-85.

48. Brentnall TA, Crispin DA, Rabinovitch PS, Haggitt RC, Rubin CE, Stevens AC, et al. Mutations in the p53 gene: an early marker of neoplastic progression in ulcerative colitis. Gastroenterology. 1994;107(2):369-78.

49. Hussain SP, Amstad P, Raja K, Ambs S, Nagashima M, Bennett WP, et al. Increased p53 mutation load in noncancerous colon tissue from ulcerative colitis: a cancer-prone chronic inflammatory disease. Cancer research. 2000;60(13):3333-7.

50. Yin J, Harpaz N, Tong Y, Huang Y, Laurin J, Greenwald BD, et al. p53 point mutations in dysplastic and cancerous ulcerative colitis lesions. Gastroenterology. 1993;104(6):1633-9.

51. Azarschab P, Porschen R, Gregor M, Blin N, Holzmann K. Epigenetic control of the E-cadherin gene (CDH1) by CpG methylation in colectomy samples of patients with ulcerative colitis. Genes, Chromosomes and Cancer. 2002;35(2):121-6.

52. Azuara D, Rodriguez-Moranta F, de Oca J, Sanjuan X, Guardiola J, Lobaton T, et al. Novel methylation panel for the early detection of neoplasia in high-risk ulcerative colitis and Crohn's colitis patients. Inflamm Bowel Dis. 2013;19(1):165-73.

53. Bai X, Zhu Y, Pu W, Xiao L, Li K, Xing C, et al. Circulating DNA and its methylation level in inflammatory bowel disease and related colon cancer. International journal of clinical and experimental pathology. 2015;8(10):13764-9.

54. Barnicle A, Seoighe C, Greally JM, Golden A, Egan LJ. Inflammation-associated DNA methylation patterns in epithelium of ulcerative colitis. Epigenetics. 2017;12(8):591-606.

55. Beggs AD, James J, Caldwell G, Prout T, Dilworth MP, Taniere P, et al. Discovery and Validation of Methylation Biomarkers for Ulcerative Colitis Associated Neoplasia. Inflamm Bowel Dis. 2018;24(7):1503-9.

56. Cooke J, Zhang H, Greger L, Silva A-L, Massey D, Dawson C, et al. Mucosal genome-wide methylation changes in inflammatory bowel disease. Inflammatory bowel diseases. 2012;18(11):2128-37.

57. Emmett RA, Davidson KL, Gould NJ, Arasaradnam RP. DNA methylation patterns in ulcerative colitis-associated cancer: a systematic review. Epigenomics. 2017;9(7):1029-42.

58. Garrity-Park MM, Loftus EV, Sandborn WJ, Bryant SC, Smyrk TC. Methylation Status of Genes in Non-Neoplastic Mucosa From Patients With Ulcerative Colitis-Associated Colorectal Cancer. The American Journal of Gastroenterology. 2010;105(7):1610-9.

59. Hasler R, Feng Z, Backdahl L, Spehlmann ME, Franke A, Teschendorff A, et al. A functional methylome map of ulcerative colitis. Genome Res. 2012;22(11):2130-7.

60. Maeda O, Ando T, Watanabe O, Ishiguro K, Ohmiya N, Niwa Y, et al. DNA hypermethylation in colorectal neoplasms and inflammatory bowel disease: a mini review. Inflammopharmacology. 2006;14(5-6):204-6.

61. Olaru AV, Cheng Y, Agarwal R, Yang J, David S, Abraham JM, et al. Unique patterns of CpG island methylation in inflammatory bowel disease-associated colorectal cancers. Inflammatory Bowel Diseases. 2012;18(4):641-8.

62. Ueda Y, Ando T, Nanjo S, Ushijima T, Sugiyama T. DNA methylation of microRNA-124a is a potential risk marker of colitis-associated cancer in patients with ulcerative colitis. Dig Dis Sci. 2014;59(10):2444-51.

63. Ullman TA, Itzkowitz SH. Intestinal Inflammation and Cancer. Gastroenterology. 2011;140(6):1807-16.e1.

64. Zhao W, Qi L, Qin Y, Wang H, Chen B, Wang R, et al. Functional Comparison between Genes Dysregulated in Ulcerative Colitis and Colorectal Carcinoma. PLoS ONE. 2013;8(8):e71989-e.

65. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. Br J Surg. 2002;89(7):845-60.

66. Rhodes JM, Campbell BJ. Inflammation and colorectal cancer: IBD-associated and sporadic cancer compared. Trends in molecular medicine. 2002;8(1):10-6.

67. Baker KT, Salk JJ, Brentnall TA, Risques RA. Precancer in ulcerative colitis: the role of the field effect and its clinical implications. Carcinogenesis. 2018;39(1):11-20.

68. Lim CH, Dixon MF, Vail A, Forman D, Lynch DA, Axon AT. Ten year follow up of ulcerative colitis patients with and without low grade dysplasia. Gut. 2003;52(8):1127-32.

69. Younus N, Abid M, Hashmi AA, Aijaz S, Edhi MM, Sheikh AK, et al. Colorectal dysplasia and adenocarcinoma in patients with ulcerative colitis: an experience from a tertiary care hospital. World J Surg Oncol. 2018;16(1):81.

70. Bai J, Li Y, Shao T, Zhao Z, Wang Y, Wu A, et al. Integrating analysis reveals microRNA-mediated pathway crosstalk among Crohn's disease, ulcerative colitis and colorectal cancer. Molecular BioSystems. 2014;10(9):2317-.

71. Eaden JA, Mayberry JF. Guidelines for screening and surveillance of asymptomatic colorectal cancer in patients with inflammatory bowel disease. Gut. 2002;51:V10-V2.

72. Roessner A, Kuester D, Malfertheiner P, Schneider-Stock R. Oxidative stress in ulcerative colitis-associated carcinogenesis. Pathol Res Pract. 2008;204(7):511-24.

73. Jena G, Trivedi PP, Sandala B. Oxidative stress in ulcerative colitis: an old concept but a new concern. Free Radic Res. 2012;46(11):1339-45.

74. Willenbucher RF, Aust DE, Chang CG, Zelman SJ, Ferrell LD, Moore DH, et al. Genomic instability is an early event during the progression pathway of ulcerative-colitis-related neoplasia. The American journal of pathology. 1999;154(6):1825-30.

75. Yaeger R, Shah MA, Miller VA, Kelsen JR, Wang K, Heins ZJ, et al. Genomic Alterations Observed in Colitis-Associated Cancers Are Distinct From Those Found in Sporadic Colorectal Cancers and Vary by Type of Inflammatory Bowel Disease. Gastroenterology. 2016;151(2):278-87 e6.

76. Chandrasinghe P, Cereser B, Moorghen M, Al Bakir I, Tabassum N, Hart A, et al. Role of SMAD proteins in colitis-associated cancer: From known to the unknown. Oncogene. 2018;37(1).

77. Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat Rev Cancer. 2017;17(4):268.

78. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350-6.

79. Thanki K, Nicholls ME, Gajjar A, Senagore AJ, Qiu S, Szabo C, et al. Consensus Molecular Subtypes of Colorectal Cancer and their Clinical Implications. Int Biol Biomed J. 2017;3(3):105-11.

80. Fujiwara I, Yashiro M, Kubo N, Maeda K, Hirakawa K. Ulcerative colitis-associated colorectal cancer is frequently associated with the microsatellite instability pathway. Dis Colon Rectum. 2008;51(9):1387-94.

81. Lovisa S, Genovese G, Danese S. Role of Epithelial-to-Mesenchymal Transition in Inflammatory Bowel Disease. J Crohns Colitis. 2019;13(5):659-68.

82. Yashiro M. Ulcerative colitis-associated colorectal cancer. World J Gastroenterol. 2014;20(44):16389-97.

83. Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G, et al. Immune evasion before tumour invasion in early lung squamous carcinogenesis. Nature. 2019;571(7766):570-5.

84. Al Bakir I, Curtius K, Graham TA. From Colitis to Cancer: An Evolutionary Trajectory That Merges Maths and Biology. Front Immunol. 2018;9:2368.

85. Baker AM, Cross W, Curtius K, Al Bakir I, Choi CR, Davis HL, et al. Evolutionary history of human colitis-associated colorectal cancer. Gut. 2019;68(6):985-95.

86. Choi CR, Al Bakir I, Ding NJ, Lee GH, Askari A, Warusavitarne J, et al. Cumulative burden of inflammation predicts colorectal neoplasia risk in ulcerative colitis: a large single-centre study. Gut. 2019;68(3):414-22.

87. Watanabe T, Konishi T, Kishimoto J, Kotake K, Muto T, Sugihara K, et al. Ulcerative colitis-associated colorectal cancer shows a poorer survival than sporadic colorectal cancer: a nationwide Japanese study. Inflamm Bowel Dis. 2011;17(3):802-8.

88. Higashi D, Futami K, Ishibashi Y, Egawa Y, Maekawa T, Matsui T, et al. Clinical course of colorectal cancer in patients with ulcerative colitis. Anticancer Res. 2011;31(7):2499-504.

89. Han YD, Al Bandar MH, Dulskas A, Cho MS, Hur H, Min BS, et al. Prognosis of ulcerative colitis colorectal cancer vs. sporadic colorectal cancer: propensity score matching analysis. BMC Surg. 2017;17(1):28.

90. van Heerden JA, Beart RW, Jr. Carcinoma of the colon and rectum complicating chronic ulcerative colitis. Dis Colon Rectum. 1980;23(3):155-9.

91. Lavery IC, Chiulli RA, Jagelman DG, Fazio VW, Weakley FL. Survival with carcinoma arising in mucosal ulcerative colitis. Ann Surg. 1982;195(4):508-12.

92. Aarnio M, Mustonen H, Mecklin JP, Jarvinen HJ. Prognosis of colorectal cancer varies in different high-risk conditions. Ann Med. 1998;30(1):75-80.

93. Jensen AB, Larsen M, Gislum M, Skriver MV, Jepsen P, Norgaard B, et al. Survival after colorectal cancer in patients with ulcerative colitis: a nationwide population-based Danish study. Am J Gastroenterol. 2006;101(6):1283-7.

94. Dugum M, Lin J, Lopez R, Estfan B, Manilich E, Stocchi L, et al. Recurrence and survival rates of inflammatory bowel disease-associated colorectal cancer following postoperative chemotherapy: a comparative study. Gastroenterol Rep (Oxf). 2017;5(1):57-61.

95. Leowardi C, Schneider ML, Hinz U, Harnoss JM, Tarantino I, Lasitschka F, et al. Prognosis of Ulcerative Colitis-Associated Colorectal Carcinoma Compared to

Sporadic Colorectal Carcinoma: A Matched Pair Analysis. Ann Surg Oncol. 2016;23(3):870-6.

96. Bogach J, Pond G, Eskicioglu C, Seow H. Age-Related Survival Differences in Patients With Inflammatory Bowel Disease-Associated Colorectal Cancer: A Population-Based Cohort Study. Inflamm Bowel Dis. 2019;25(12):1957-65.

97. Winther KV, Jess T, Langholz E, Munkholm P, Binder V. Survival and cause-specific mortality in ulcerative colitis: follow-up of a population-based cohort in Copenhagen County. Gastroenterology. 2003;125(6):1576-82.

98. Lutgens MW, Oldenburg B, Siersema PD, van Bodegraven AA, Dijkstra G, Hommes DW, et al. Colonoscopic surveillance improves survival after colorectal cancer diagnosis in inflammatory bowel disease. Br J Cancer. 2009;101(10):1671-5.

99. Ananthakrishnan AN, Cagan A, Cai T, Gainer VS, Shaw SY, Churchill S, et al. Colonoscopy is associated with a reduced risk for colon cancer and mortality in patients with inflammatory bowel diseases. Clin Gastroenterol Hepatol. 2015;13(2):322-9 e1.

100. Hata K, Watanabe T, Kazama S, Suzuki K, Shinozaki M, Yokoyama T, et al. Earlier surveillance colonoscopy programme improves survival in patients with ulcerative colitis associated colorectal cancer: results of a 23-year surveillance programme in the Japanese population. Br J Cancer. 2003;89(7):1232-6.

101. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38(1):23-38.

102. Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? Genes Cancer. 2011;2(6):607-17.

103. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007;450(7171):908-12.

104. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol. 2005;45:629-56.

105. Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, Brown CJ. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum Mol Genet. 2015;24(6):1528-39.

106. Sharp AJ, Stathaki E, Migliavacca E, Brahmachary M, Montgomery SB, Dupre Y, et al. DNA methylation profiles of human active and inactive X chromosomes. Genome Res. 2011;21(10):1592-600.

107. Xin Y, O'Donnell AH, Ge Y, Chanrion B, Milekic M, Rosoklija G, et al. Role of CpG context and content in evolutionary signatures of brain DNA methylation. Epigenetics. 2011;6(11):1308-18.

108. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet. 2009;10(5):295-304.

109. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465-76.

110. Pacis A, Mailhot-Leonard F, Tailleux L, Randolph HE, Yotova V, Dumaine A, et al. Gene activation precedes DNA demethylation in response to infection in human dendritic cells. Proc Natl Acad Sci U S A. 2019;116(14):6938-43.

111. Chakraborty A, Viswanathan P. Methylation-Demethylation Dynamics: Implications of Changes in Acute Kidney Injury. Anal Cell Pathol (Amst). 2018;2018:8764384.

112. Shi Y, Massague J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell. 2003;113(6):685-700.

113. Slieker RC, Roost MS, van Iperen L, Suchiman HE, Tobi EW, Carlotti F, et al. DNA Methylation Landscapes of Human Fetal Development. PLoS Genet. 2015;11(10):e1005583.

114. Watt F, Molloy PL. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes & Development. 1988;2(9):1136-43.

115. Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell. 1989;58(3):499-507.

116. Cardinale CJ, Wei Z, Li J, Zhu J, Gu M, Baldassano RN, et al. Transcriptome profiling of human ulcerative colitis mucosa reveals altered expression of pathways enriched in genetic susceptibility loci. PLoS One. 2014;9(5):e96153.

117. Yu W, Lin Z, Hegarty JP, Chen X, Kelly AA, Wang Y, et al. Genes differentially regulated by NKX2-3 in B cells between ulcerative colitis and Crohn's disease patients and possible involvement of EGR1. Inflammation. 2012;35(3):889-99.

118. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010;42(12):1118-25.

119. Kim SW, Kim ES, Moon CM, Kim TI, Kim WH, Cheon JH. Abnormal genetic and epigenetic changes in signal transducer and activator of transcription 4 in the pathogenesis of inflammatory bowel diseases. Digestive diseases and sciences. 2012;57(10):2600-7.

120. Ventham NT, Kennedy NA, Adams AT, Kalla R, Heath S, O'Leary KR, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. Nat Commun. 2016;7:13507.

121. McDermott E, Ryan EJ, Tosetto M, Gibson D, Burrage J, Keegan D, et al. DNA Methylation Profiling in Inflammatory Bowel Disease Provides New Insights into Disease Pathogenesis. Journal of Crohn's & colitis. 2016;10(1):77-86.

122. Balic M, Pichler M, Strutz J, Heitzer E, Ausch C, Samonigg H, et al. High quality assessment of DNA methylation in archival tissues from colorectal cancer patients using quantitative high-resolution melting analysis. J Mol Diagn. 2009;11(2):102-8.

123. Witte T, Plass C, Gerhauser C. Pan-cancer patterns of DNA methylation. Genome Med. 2014;6(8):66.

124. Boirivant M, Pallone F, Di Giacinto C, Fina D, Monteleone I, Marinaro M, et al. Inhibition of Smad7 With a Specific Antisense Oligonucleotide Facilitates TGF-β1–Mediated Suppression of Colitis. Gastroenterology. 2006;131(6):1786-98.

125. Fantini MC, Rizzo A, Fina D, Caruso R, Sarra M, Stolfi C, et al. Smad7 expression in T cells protects from colitis-associated colorectal cancer. Gastroenterology. 2009;136(5):A8-A9.

126. Marafini I, Sedda S, Fusco DD, Figliuzzi MM, Pallone F, Monteleone G. Smad7 Sustains Inflammation in the Gut: From Bench to Bedside.

127. Monteleone G, Fantini MC, Onali S, Zorzi F, Sancesario G, Bernardini S, et al. Phase I Clinical Trial of Smad7 Knockdown Using Antisense Oligonucleotide in Patients With Active Crohn's Disease. Molecular Therapy. 2012;20(4):870-6.

128. Monteleone G, Neurath MF, Ardizzone S, Di Sabatino A, Fantini MC, Castiglione F, et al. Mongersen, an Oral SMAD7 Antisense Oligonucleotide, and Crohn's Disease. New England Journal of Medicine. 2015;372(12):1104-13.

129. Sekelsky JJ, Newfeld SJ, Raftery LA, Chartoff EH, Gelbart WM. Genetic characterization and cloning of mothers against dpp, a gene required for decapentaplegic function in Drosophila melanogaster. Genetics. 1995;139(3).

130. Savage C, Das P, Finelli AL, Townsend SR, Sun CY, Baird SE, et al. Caenorhabditis elegans genes sma-2, sma-3, and sma-4 define a conserved family of transforming growth factor beta pathway components. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(2):790-4.

131. Shi Y, Massagué J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell. 2003;113(6):685-700.

132. Gomes Fernandes M, Dries R, Roost Matthias S, Semrau S, de Melo Bernardo A, Davis Richard P, et al. BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells. Stem Cell Reports. 2016;6(1):85-94.

133. Papageorgis P, Lambert AW, Ozturk S, Gao F, Pan H, Manne U, et al. Smad signaling is required to maintain epigenetic silencing during breast cancer progression. Cancer research. 2010;70(3):968-78.

134. Mullen AC, Orlando DA, Newman JJ, Loven J, Kumar RM, Bilodeau S, et al. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell. 2011;147(3):565-76.

135. Lu Y, Wang L, Li H, Li Y, Ruan Y, Lin D, et al. SMAD2 Inactivation Inhibits CLDN6 Methylation to Suppress Migration and Invasion of Breast Cancer Cells. Int J Mol Sci. 2017;18(9).

136. Thillainadesan G, Chitilian JM, Isovic M, Ablack JN, Mymryk JS, Tini M, et al. TGF-beta-dependent active demethylation and expression of the p15ink4b tumor suppressor are impaired by the ZNF217/CoREST complex. Mol Cell. 2012;46(5):636-49.

137. Guo W, Zhang M, Shen S, Guo Y, Kuang G, Yang Z, et al. Aberrant methylation and decreased expression of the TGF-beta/Smad target gene FBXO32 in esophageal squamous cell carcinoma. Cancer. 2014;120(16):2412-23.

138. Chou J-L, Chen L-Y, Lai H-C, Chan MWY. TGF-β: friend or foe? The role of TGF-β/SMAD signaling in epigenetic silencing of ovarian cancer and its implication in epigenetic therapy. Expert Opinion on Therapeutic Targets. 2010;14(11):1213-23.

139. Bai J, Xi Q. Crosstalk between TGF-beta signaling and epigenome. Acta Biochim Biophys Sin (Shanghai). 2018;50(1):60-7.

140. Monteleone G, Caruso R, Pallone F. Role of Smad7 in inflammatory bowel diseases. World journal of gastroenterology. 2012;18(40):5664-8.

141. Monteleone G, Del Vecchio Blanco G, Monteleone I, Fina D, Caruso R, Gioia V, et al. Post-transcriptional regulation of Smad7 in the gut of patients with inflammatory bowel disease. Gastroenterology. 2005;129(5):1420-9.

142. Monteleone G, Kumberova A, Croft NM, McKenzie C, Steer HW, MacDonald TT. Blocking Smad7 restores TGF-beta1 signaling in chronic inflammatory bowel disease. The Journal of clinical investigation. 2001;108(4):601-9.

143. Sands BE, Feagan BG, Sandborn WJ, Schreiber S, Peyrin-Biroulet L, Frederic Colombel J, et al. Mongersen (GED-0301) for Active Crohn's Disease: Results of a Phase 3 Study. Am J Gastroenterol. 2020;115(5):738-45.

144. Fantini MC, Rizzo A, Fina D, Caruso R, Sarra M, Stolfi C, et al. Smad7 Controls Resistance of Colitogenic T Cells to Regulatory T Cell-Mediated Suppression. Gastroenterology. 2009;136(4):1308-16.e3.

145. Stolfi C, Marafini I, De Simone V, Pallone F, Monteleone G. The dual role of Smad7 in the control of cancer growth and metastasis. 2013. p. 23774-90.

146. Rizzo A, Waldner MJ, Stolfi C, Sarra M, Fina D, Becker C, et al. Smad7 Expression in T cells Prevents Colitis-Associated Cancer. Cancer Research. 2011;71(24).

147. Becker C, Fantini MC, Schramm C, Lehr HA, Wirtz S, Nikolaev A, et al. TGF-beta suppresses tumor progression in colon cancer by inhibition of IL-6 trans-signaling. Immunity. 2004;21(4):491-501.

148. Chaudhury A, Howe PH. The tale of transforming growth factor-beta (TGFβ) signaling: A soigné enigma. IUBMB Life. 2009;61(10):929-39.

149. Massagué J. TGFβ in Cancer. Cell. 2008;134(2):215-30.

150. Principe DR, Doll JA, Bauer J, Jung B, Munshi HG, Bartholin L, et al. TGF-β: duality of function between tumor prevention and carcinogenesis. Journal of the National Cancer Institute. 2014;106(2):djt369-djt.

151. Wakefield LM, Hill CS. Beyond TGFβ: roles of other TGFβ superfamily members in cancer. Nature Reviews Cancer. 2013;13(5):328-41.

152. Wirtz S, Popp V, Kindermann M, Gerlach K, Weigmann B, Fichtner-Feigl S, et al. Chemically induced mouse models of acute and chronic intestinal inflammation. Nat Protoc. 2017;12(7):1295-309.

153. Hernandez-Chirlaque C, Aranda CJ, Ocon B, Capitan-Canadas F, Ortega-Gonzalez M, Carrero JJ, et al. Germ-free and Antibiotic-treated Mice are Highly Susceptible to Epithelial Injury in DSS Colitis. J Crohns Colitis. 2016;10(11):1324-35.

154. Antoniou E, Margonis GA, Angelou A, Pikouli A, Argiri P, Karavokyros I, et al. The TNBS-induced colitis animal model: An overview. Ann Med Surg (Lond). 2016;11:9-15.

155. Boirivant M, Fuss IJ, Chu A, Strober W. Oxazolone colitis: A murine model of T helper cell type 2 colitis treatable with antibodies to interleukin 4. J Exp Med. 1998;188(10):1929-39.

156. MacPherson BR, Pfeiffer CJ. Experimental production of diffuse colitis in rats. Digestion. 1978;17(2):135-50.

157. Fenga C, Gangemi S, Costa C. Benzene exposure is associated with epigenetic changes (Review). Mol Med Rep. 2016;13(4):3401-5.

158. Hiraku Y, Kawanishi S. Oxidative DNA damage and apoptosis induced by benzene metabolites. Cancer Res. 1996;56(22):5172-8.

159. Kim M, Bae M, Na H, Yang M. Environmental toxicants--induced epigenetic alterations and their reversers. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 2012;30(4):323-67.

160. Keubler LM, Buettner M, Hager C, Bleich A. A Multihit Model: Colitis Lessons from the Interleukin-10-deficient Mouse. Inflamm Bowel Dis. 2015;21(8):1967-75.

161. Wilk JN, Bilsborough J, Viney JL. The mdr1a-/- mouse model of spontaneous colitis: a relevant and appropriate animal model to study inflammatory bowel disease. Immunol Res. 2005;31(2):151-9.

162. Spalinger MR, Sayoc-Becerra A, Santos AN, Shawki A, Canale V, Krishnan M, et al. PTPN2 Regulates Interactions Between Macrophages and Intestinal Epithelial Cells to Promote Intestinal Barrier Function. Gastroenterology. 2020;159(5):1763-77 e14.

163. Prattis S, Jurjus A. Spontaneous and transgenic rodent models of inflammatory bowel disease. Lab Anim Res. 2015;31(2):47-68.

164. Hahm KB, Im YH, Parks TW, Park SH, Markowitz S, Jung HY, et al. Loss of transforming growth factor beta signalling in the intestine contributes to tissue injury in inflammatory bowel disease. Gut. 2001;49(2):190-8.

165. Snider AJ, Bialkowska AB, Ghaleb AM, Yang VW, Obeid LM, Hannun YA. Murine Model for Colitis-Associated Cancer of the Colon. Methods Mol Biol. 2016;1438:245-54.

166. Kanneganti M, Mino-Kenudson M, Mizoguchi E. Animal models of colitis-associated carcinogenesis. J Biomed Biotechnol. 2011;2011:342637.

167. Greten FR, Eckmann L, Greten TF, Park JM, Li ZW, Egan LJ, et al. IKKbeta links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. Cell. 2004;118(3):285-96.

168. Paradisi A, Maisse C, Coissieux MM, Gadot N, Lepinasse F, Delloye-Bourgeois C, et al. Netrin-1 up-regulation in inflammatory bowel diseases is required for colorectal cancer progression. Proc Natl Acad Sci U S A. 2009;106(40):17146-51.

169. Grivennikov S, Karin E, Terzic J, Mucida D, Yu GY, Vallabhapurapu S, et al. IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. Cancer Cell. 2009;15(2):103-13.

170. Tanaka T, Kohno H, Suzuki R, Yamada Y, Sugie S, Mori H. A novel inflammation-related mouse colon carcinogenesis model induced by azoxymethane and dextran sodium sulfate. Cancer Sci. 2003;94(11):965-73.

171. Tanaka T. Colorectal carcinogenesis: Review of human and experimental animal studies. J Carcinog. 2009;8:5.

172. Slattery ML, Mullany LE, Sakoda L, Samowitz WS, Wolff RK, Stevens JR, et al. The NF-kappaB signalling pathway in colorectal cancer: associations between dysregulated gene and miRNA expression. J Cancer Res Clin Oncol. 2018;144(2):269-83.

173. Slattery ML, Lundgreen A, Kadlubar SA, Bondurant KL, Wolff RK. JAK/STAT/SOCS-signaling pathway and colon and rectal cancer. Mol Carcinog. 2013;52(2):155-66.

174. Angus HCK, Butt AG, Schultz M, Kemp RA. Intestinal Organoids as a Tool for Inflammatory Bowel Disease Research. Front Med (Lausanne). 2019;6:334.

175. Ray G, Longworth MS. Epigenetics, DNA Organization, and Inflammatory Bowel Disease. Inflamm Bowel Dis. 2019;25(2):235-47.

176. Liu Y, Peng J, Sun T, Li N, Zhang L, Ren J, et al. Epithelial EZH2 serves as an epigenetic determinant in experimental colitis by inhibiting TNFalpha-mediated inflammation and apoptosis. Proc Natl Acad Sci U S A. 2017;114(19):E3796-E805.

177. Turgeon N, Gagne JM, Blais M, Gendron FP, Boudreau F, Asselin C. The acetylome regulators Hdac1 and Hdac2 differently modulate intestinal epithelial cell dependent homeostatic responses in experimental colitis. Am J Physiol Gastrointest Liver Physiol. 2014;306(7):G594-605.

178. Sakatani T, Kaneda A, Iacobuzio-Donahue CA, Carter MG, de Boom Witzel S, Okano H, et al. Loss of imprinting of Igf2 alters intestinal maturation and tumorigenesis in mice. Science. 2005;307(5717):1976-8.

179. Small CL, Reid-Yu SA, McPhee JB, Coombes BK. Persistent infection with Crohn's disease-associated adherent-invasive Escherichia coli leads to chronic inflammation and intestinal fibrosis. Nat Commun. 2013;4:1957.

180. Liu Z, Cao W, Xu L, Chen X, Zhan Y, Yang Q, et al. The histone H3 lysine-27 demethylase Jmjd3 plays a critical role in specific regulation of Th17 cell differentiation. J Mol Cell Biol. 2015;7(6):505-16.

181. Antignano F, Burrows K, Hughes MR, Han JM, Kron KJ, Penrod NM, et al. Methyltransferase G9A regulates T cell differentiation during murine intestinal inflammation. J Clin Invest. 2014;124(5):1945-55.

182. Ishiguro K, Watanabe O, Nakamura M, Yamamura T, Matsushita M, Goto H, et al. Inhibition of KDM4A activity as a strategy to suppress interleukin-6 production and attenuate colitis induction. Clin Immunol. 2017;180:120-7.

183. Papadakis KA, Krempski J, Svingen P, Xiong Y, Sarmento OF, Lomberk GA, et al. Kruppel-like factor KLF10 deficiency predisposes to colitis through colonic macrophage dysregulation. Am J Physiol Gastrointest Liver Physiol. 2015;309(11):G900-9.

184. Weiser M, Simon JM, Kochar B, Tovar A, Israel JW, Robinson A, et al. Molecular classification of Crohn's disease reveals two clinically relevant subtypes. Gut. 2018;67(1):36-42.

185. Tehranchi AK, Myrthil M, Martin T, Hie BL, Golan D, Fraser HB. Pooled ChIP-Seq Links Variation in Transcription Factor Binding to Complex Disease Risk. Cell. 2016;165(3):730-41.

186. Simard ML, Mourier A, Greaves LC, Taylor RW, Stewart JB. A novel histochemistry assay to assess and quantify focal cytochrome c oxidase deficiency. J Pathol. 2018;245(3):311-23.

187. van Beers EH, Joosse SA, Ligtenberg MJ, Fles R, Hogervorst FBL, Verhoef S, et al. A multiplex PCR predictor for aCGH success of FFPE samples. British journal of cancer. 2006;94(2):333-7.

188. Scott JE. On the mechanism of the methyl green-pyronin stain for nucleic acids. Histochemie. 1967;9(1):30-47.

189. Ziller MJ, Hansen KD, Meissner A, Aryee MJ. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Methods. 2015;12(3):230-2, 1 p following 2.

190. Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, et al. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. Nat Biotechnol. 2010;28(5):516-20.

191. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315-22.

192. van Montfoort AP, Hanssen LL, de Sutter P, Viville S, Geraedts JP, de Boer P. Assisted reproduction treatment and epigenetic inheritance. Hum Reprod Update. 2012;18(2):171-97.

193. Novakovic B, Lewis S, Halliday J, Kennedy J, Burgner DP, Czajko A, et al. Assisted reproductive technologies are associated with limited epigenetic variation at birth that largely resolves by adulthood. Nat Commun. 2019;10(1):3922.

194. Osier MV, Zhao H, Cheung K-H. Handling multiple testing while interpreting microarrays with the Gene Ontology Database. BMC Bioinformatics. 2004;5(1):124.

195. Chen J, Aronow BJ, Jegga AG. Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics. 2009;10(1):73.

196. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Research. 2009;37(suppl_2):W305-W11.

197. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11:587.

198. Grade M, Hummon AB, Camps J, Emons G, Spitzner M, Gaedcke J, et al. A genomic strategy for the functional validation of colorectal cancer genes identifies potential therapeutic targets. Int J Cancer. 2011;128(5):1069-79.

199. Shvartsur A, Bonavida B. Trop2 and its overexpression in cancers: regulation and clinical/therapeutic implications. Genes Cancer. 2015;6(3-4):84-105.

200. Benelli D, Cialfi S, Pinzaglia M, Talora C, Londei P. The translation factor eIF6 is a Notch-dependent regulator of cell migration and invasion. PLoS One. 2012;7(2):e32047.

201. Lin J, Yu X, Xie L, Wang P, Li T, Xiao Y, et al. eIF6 Promotes Colorectal Cancer Proliferation and Invasion by Regulating AKT-Related Signaling Pathways. J Biomed Nanotechnol. 2019;15(7):1556-67.

202. Zhu W, Li GX, Chen HL, Liu XY. The role of eukaryotic translation initiation factor 6 in tumors. Oncol Lett. 2017;14(1):3-9.

203. Chu G, Li J, Zhao Y, Liu N, Zhu X, Liu Q, et al. Identification and verification of PRDX1 as an inflammation marker for colorectal cancer progression. Am J Transl Res. 2016;8(2):842-59.

204. Ding C, Fan X, Wu G. Peroxiredoxin 1 - an antioxidant enzyme in cancer. J Cell Mol Med. 2017;21(1):193-202.

205. Li HX, Sun XY, Yang SM, Wang Q, Wang ZY. Peroxiredoxin 1 promoted tumor metastasis and angiogenesis in colorectal cancer. Pathol Res Pract. 2018;214(5):655-60.

206. Liu J, Zhu H, Zhong N, Jiang Z, Xu L, Deng Y, et al. Gene silencing of USP1 by lentivirus effectively inhibits proliferation and invasion of human osteosarcoma cells. Int J Oncol. 2016;49(6):2549-57.

207. Ma A, Tang M, Zhang L, Wang B, Yang Z, Liu Y, et al. USP1 inhibition destabilizes KPNA2 and suppresses breast cancer metastasis. Oncogene. 2019;38(13):2405-19.

208. Xu X, Li S, Cui X, Han K, Wang J, Hou X, et al. Inhibition of Ubiquitin Specific Protease 1 Sensitizes Colorectal Cancer Cells to DNA-Damaging Chemotherapeutics. Front Oncol. 2019;9:1406.

209. Xu D, Zhang H, Wang X, Chen Y. Expression of IRAK3 is associated with colitisassociated tumorigenesis in mice. Mol Med Rep. 2017;16(3):3415-20.

210. Cantini L, Isella C, Petti C, Picco G, Chiola S, Ficarra E, et al. MicroRNA–mRNA interactions underlying colorectal cancer molecular subtypes. Nat Commun. 2015;6:8878-.

211. Lee J, Katzenmaier EM, Kopitz J, Gebert J. Reconstitution of TGFBR2 in HCT116 colorectal cancer cells causes increased LFNG expression and enhanced N-acetyl-d-glucosamine incorporation into Notch1. Cell Signal. 2016;28(8):1105-13.

212. Reedijk M, Odorcic S, Zhang H, Chetty R, Tennert C, Dickson BC, et al. Activation of Notch signaling in human colon adenocarcinoma. Int J Oncol. 2008;33(6):1223-9.

213. Camps J, Pitt JJ, Emons G, Hummon AB, Case CM, Grade M, et al. Genetic amplification of the NOTCH modulator LNX2 upregulates the WNT/beta-catenin pathway in colorectal cancer. Cancer Res. 2013;73(6):2003-13.

214. Karl J, Wild N, Tacke M, Andres H, Garczarek U, Rollinger W, et al. Improved diagnosis of colorectal cancer using a combination of fecal occult blood and novel fecal protein markers. Clin Gastroenterol Hepatol. 2008;6(10):1122-8.

215. Wild N, Andres H, Rollinger W, Krause F, Dilba P, Tacke M, et al. A combination of serum markers for the early detection of colorectal cancer. Clin Cancer Res. 2010;16(24):6111-21.

216. Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, et al. Functions of S100 proteins. Curr Mol Med. 2013;13(1):24-57.

217. Mokarram P, Kumar K, Brim H, Naghibalhossaini F, Saberi-firoozi M, Nouraie M, et al. Distinct high-profile methylated genes in colorectal cancer. PLoS One. 2009;4(9):e7012.

218. Miyoshi H, Morishita A, Tani J, Sakamoto T, Fujita K, Katsura A, et al. Expression profiles of 507 proteins from a biotin label-based antibody array in human colorectal cancer. Oncol Rep. 2014;31(3):1277-81.

219. 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.

220. Laggai S, Kessler SM, Boettcher S, Lebrun V, Gemperlein K, Lederer E, et al. The IGF2 mRNA binding protein p62/IGF2BP2-2 induces fatty acid elongation as a critical feature of steatosis. J Lipid Res. 2014;55(6):1087-97.

221. Silva TD, Vidigal VM, Felipe AV, JM DEL, Neto RA, Saad SS, et al. DNA methylation as an epigenetic biomarker in colorectal cancer. Oncol Lett. 2013;6(6):1687-92.

222. Kasprzak A, Adamek A. Insulin-Like Growth Factor 2 (IGF2) Signaling in Colorectal Cancer-From Basic Research to Potential Clinical Applications. Int J Mol Sci. 2019;20(19).

223. Gould NJD, Davidson KLM, Nwokolo CU, Arasaradnam RP. A systematic review of the role of DNA methylation on inflammatory genes in ulcerative colitis. Epigenomics. 2016;8(5):667-84.

224. Zhang J, Lu Y, Yue X, Li H, Luo X, Wang Y, et al. MiR-124 suppresses growth of human colorectal cancer by inhibiting STAT3. PLoS One. 2013;8(8):e70300.

225. Ayatollahi H, Sadeghian MH, Naderi M, Jafarian AH, Shams SF, Motamedirad N, et al. Quantitative assessment of Wilms tumor 1 expression by real-time quantitative polymerase chain reaction in patients with acute myeloblastic leukemia. J Res Med Sci. 2017;22:54.

226. Oji Y, Yamamoto H, Nomura M, Nakano Y, Ikeba A, Nakatsuka S, et al. Overexpression of the Wilms' tumor gene WT1 in colorectal adenocarcinoma. Cancer Sci. 2003;94(8):712-7.

227. Spainhour JC, Lim HS, Yi SV, Qiu P. Correlation Patterns Between DNA Methylation and Gene Expression in The Cancer Genome Atlas. Cancer Inform. 2019;18:1176935119828776.

228. Gong YZ, Ruan GT, Liao XW, Wang XK, Liao C, Wang S, et al. Diagnostic and prognostic values of integrin alpha subfamily mRNA expression in colon adenocarcinoma. Oncol Rep. 2019;42(3):923-36.

229. Kok-Sin T, Mokhtar NM, Ali Hassan NZ, Sagap I, Mohamed Rose I, Harun R, et al. Identification of diagnostic markers in colorectal cancer via integrative epigenomics and genomics data. Oncol Rep. 2015;34(1):22-32.

230. Choi JS, Kim KH, Jeon YK, Kim SH, Jang SG, Ku JL, et al. Promoter hypermethylation of the ADAM23 gene in colorectal cancer cell lines and cancer tissues. Int J Cancer. 2009;124(6):1258-62.

231. Kalinkova L, Zmetakova I, Smolkova B, Minarik G, Sedlackova T, Horvathova Kajabova V, et al. Decreased methylation in the SNAI2 and ADAM23 genes associated with de-differentiation and haematogenous dissemination in breast cancers. BMC Cancer. 2018;18(1):875.

232. Takada H, Imoto I, Tsuda H, Nakanishi Y, Ichikura T, Mochizuki H, et al. ADAM23, a possible tumor suppressor gene, is frequently silenced in gastric cancers by homozygous deletion or aberrant promoter hypermethylation. Oncogene. 2005;24(54):8051-60.

233. Zhao J, Li P, Feng H, Wang P, Zong Y, Ma J, et al. Cadherin-12 contributes to tumorigenicity in colorectal cancer by promoting migration, invasion, adhersion and angiogenesis. J Transl Med. 2013;11:288.

234. Ma J, Zhao J, Lu J, Wang P, Feng H, Zong Y, et al. Cadherin-12 enhances proliferation in colorectal cancer cells and increases progression by promoting EMT. Tumour Biol. 2016;37(7):9077-88.

235. Li MZ, Wang JJ, Yang SB, Li WF, Xiao LB, He YL, et al. ZEB2 promotes tumor metastasis and correlates with poor prognosis of human colorectal cancer. Am J Transl Res. 2017;9(6):2838-51.

236. Sreekumar R, Harris S, Moutasim K, DeMateos R, Patel A, Emo K, et al. Assessment of Nuclear ZEB2 as a Biomarker for Colorectal Cancer Outcome and TNM Risk Stratification. JAMA Netw Open. 2018;1(6):e183115.

237. Vu JV, De Roo A, Hardiman KM. Can ZEB2 Be Used as a Molecular Marker for Risk Stratification of Patients With Colorectal Cancer? JAMA Netw Open. 2018;1(6):e183133.

238. Lin SH, Raju GS, Huff C, Ye Y, Gu J, Chen JS, et al. The somatic mutation landscape of premalignant colorectal adenoma. Gut. 2018;67(7):1299-305.

239. Villacis RA, Miranda PM, Gomy I, Santos EM, Carraro DM, Achatz MI, et al. Contribution of rare germline copy number variations and common susceptibility loci in Lynch syndrome patients negative for mutations in the mismatch repair genes. Int J Cancer. 2016;138(8):1928-35.

240. Lee M, Chun SM, Sung CO, Kim SY, Kim TW, Jang SJ, et al. Clinical Utility of a Fully Automated Microsatellite Instability Test with Minimal Hands-on Time. J Pathol Transl Med. 2019;53(6):386-92.

241. Gu C, Cai J, Xu Z, Zhou S, Ye L, Yan Q, et al. MiR-532-3p suppresses colorectal cancer progression by disrupting the ETS1/TGM2 axis-mediated Wnt/beta-catenin signaling. Cell Death Dis. 2019;10(10):739.

242. Nakayama T, Ito M, Ohtsuru A, Naito S, Sekine I. Expression of the ets-1 proto-oncogene in human colorectal carcinoma. Mod Pathol. 2001;14(5):415-22.

243. Li S, Wu X, Xu Y, Wu S, Li Z, Chen R, et al. miR-145 suppresses colorectal cancer cell migration and invasion by targeting an ETS-related gene. Oncol Rep. 2016;36(4):1917-26.

244. Lin K, Taylor JR, Jr., Wu TD, Gutierrez J, Elliott JM, Vernes JM, et al. TMEFF2 is a PDGF-AA binding protein with methylation-associated gene silencing in multiple cancer types including glioma. PLoS One. 2011;6(4):e18608.

245. Belshaw NJ, Elliott GO, Williams EA, Bradburn DM, Mills SJ, Mathers JC, et al. Use of DNA from human stools to detect aberrant CpG island methylation of

genes implicated in colorectal cancer. Cancer Epidemiol Biomarkers Prev. 2004;13(9):1495-501.

246. Saito S, Kato J, Hiraoka S, Horii J, Suzuki H, Higashi R, et al. DNA methylation of colon mucosa in ulcerative colitis patients: Correlation with inflammatory status. Inflammatory Bowel Diseases. 2011;17(9):1955-65.

247. Li Y, Li J, Liu H, Liu Y, Cui B. Expression of MYSM1 is associated with tumor progression in colorectal cancer. PLoS One. 2017;12(5):e0177235.

248. Hlavata I, Mohelnikova-Duchonova B, Vaclavikova R, Liska V, Pitule P, Novak P, et al. The role of ABC transporters in progression and clinical outcome of colorectal cancer. Mutagenesis. 2012;27(2):187-96.

249. Kaneuchi M, Sasaki M, Tanaka Y, Shiina H, Yamada H, Yamamoto R, et al. WT1 and WT1-AS genes are inactivated by promoter methylation in ovarian clear cell adenocarcinoma. Cancer. 2005;104(9):1924-30.

250. Malik KT, Wallace JI, Ivins SM, Brown KW. Identification of an antisense WT1 promoter in intron 1: implications for WT1 gene regulation. Oncogene. 1995;11(8):1589-95.

251. Watanabe T, Kobunai T, Sakamoto E, Yamamoto Y, Konishi T, Horiuchi A, et al. Gene expression signature for recurrence in stage III colorectal cancers. Cancer. 2009;115(2):283-92.

252. Yi JK, Kim HJ, Yu DH, Park SJ, Shin MJ, Yuh HS, et al. Regulation of inflammatory responses and fibroblast-like synoviocyte apoptosis by calcineurin-binding protein 1 in mice with collagen-induced arthritis. Arthritis Rheum. 2012;64(7):2191-200.

253. Dou J, Schmidt RJ, Benke KS, Newschaffer C, Hertz-Picciotto I, Croen LA, et al. Cord blood buffy coat DNA methylation is comparable to whole cord blood methylation. Epigenetics. 2018;13(1):108-16.

254. Zheng Y, Joyce BT, Colicino E, Liu L, Zhang W, Dai Q, et al. Blood Epigenetic Age may Predict Cancer Incidence and Mortality. EBioMedicine. 2016;5:68-73.

255. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, et al. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. Age (Dordr). 2014;36(3):9648.

256. van Bemmel D, Lenz P, Liao LM, Baris D, Sternberg LR, Warner A, et al. Correlation of LINE-1 methylation levels in patient-matched buffy coat, serum, buccal cell, and bladder tumor tissue DNA samples. Cancer Epidemiol Biomarkers Prev. 2012;21(7):1143-8.

257. Sturgeon SR, Arcaro KF, Johnson MA, Balasubramanian R, Zorn M, Jerry DJ, et al. DNA methylation in paired breast epithelial and white blood cells from women undergoing reduction mammoplasty. Anticancer Res. 2014;34(6):2985-90.

258. Moarii M, Boeva V, Vert JP, Reyal F. Changes in correlation between promoter methylation and gene expression in cancer. BMC Genomics. 2015;16:873.

259. Arechederra M, Daian F, Yim A, Bazai SK, Richelme S, Dono R, et al. Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. Nat Commun. 2018;9(1):3164.

260. Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP, et al. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. Genome Biol. 2014;15(9):469.

261. Foran E, Garrity-Park MM, Mureau C, Newell J, Smyrk TC, Limburg PJ, et al. Upregulation of DNA methyltransferase-mediated gene silencing, anchorage-

independent growth, and migration of colon cancer cells by interleukin-6. Mol Cancer Res. 2010;8(4):471-81.

262. Yi JM, Kim TO. Epigenetic Alterations in Inflammatory Bowel Disease and Cancer. Intestinal Research. 2015;13(2):112-.

263. Yang ZH, Dang YQ, Ji G. Role of epigenetics in transformation of inflammation into colorectal cancer. World J Gastroenterol. 2019;25(23):2863-77.

264. Chen J, Sun H, Tang W, Zhou L, Xie X, Qu Z, et al. DNA methylation biomarkers in stool for early screening of colorectal cancer. J Cancer. 2019;10(21):5264-71.

265. Huang D, Sun W, Zhou Y, Li P, Chen F, Chen H, et al. Mutations of key driver genes in colorectal cancer progression and metastasis. Cancer Metastasis Rev. 2018;37(1):173-87.

266. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. Gastrointest Cancer Res. 2012;5(1):19-27.

267. Hartnett L, Egan LJ. Inflammation, DNA methylation and colitis-associated cancer. Carcinogenesis. 2012;33(4):723-31.

268. Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA. Accelerated age-related CpG island methylation in ulcerative colitis. Cancer research. 2001;61(9):3573-7.

269. Chen R, Lai LA, Brentnall TA, Pan S. Biomarkers for colitis-associated colorectal cancer. World journal of gastroenterology. 2016;22(35):7882-91.

270. Mojtabanezhad Shariatpanahi A, Yassi M, Nouraie M, Sahebkar A, Varshoee Tabrizi F, Kerachian MA. The importance of stool DNA methylation in colorectal cancer diagnosis: A meta-analysis. PLoS One. 2018;13(7):e0200735.

271. Leygo C, Williams M, Jin HC, Chan MWY, Chu WK, Grusch M, et al. DNA Methylation as a Noninvasive Epigenetic Biomarker for the Detection of Cancer. Dis Markers. 2017;2017:3726595.

272. Romagosa C, Simonetti S, Lopez-Vicente L, Mazo A, Lleonart ME, Castellvi J, et al. p16(Ink4a) overexpression in cancer: a tumor suppressor gene associated with senescence and high-grade tumors. Oncogene. 2011;30(18):2087-97.

273. Yoshida T, Matsumoto N, Mikami T, Okayasu I. Upregulation of p16(INK4A) and Bax in p53 wild/p53-overexpressing crypts in ulcerative colitis-associated tumours. Br J Cancer. 2004;91(6):1081-8.

274. Vincent KM, Postovit LM. A pan-cancer analysis of secreted Frizzled-related proteins: re-examining their proposed tumour suppressive function. Sci Rep. 2017;7:42719.

275. Yamamura S, Kawakami K, Hirata H, Ueno K, Saini S, Majid S, et al. Oncogenic functions of secreted Frizzled-related protein 2 in human renal cancer. Mol Cancer Ther. 2010;9(6):1680-7.

276. Chandrasinghe PCCBMMABITNHASJWJ. Role of SMAD proteins in colitis associated cancer: from known to the unknown. Oncogene.

277. De Simone V, Bevivino G, Sedda S, Izzo R, Laudisi F, Dinallo V, et al. Smad7 knockdown activates protein kinase RNA-associated eIF2? pathway leading to colon cancer cell death. Cell Death and Disease. 2017;8(3):e2681-e.

278. Stolfi C, De Simone V, Colantoni A, Franzè E, Ribichini E, Fantini MC, et al. A functional role for Smad7 in sustaining colon cancer cell growth and survival. Cell Death and Disease. 2014;5(2):e1073-e.

279. Kim S, Jacobs-Wagner C. Effects of mRNA Degradation and Site-Specific Transcriptional Pausing on Protein Expression Noise. Biophys J. 2018;114(7):1718-29.

280. Shukla A, Yang Y, Madanikia S, Ho Y, Li M, Sanchez V, et al. Elevating CLIC4 in Multiple Cell Types Reveals a TGF- Dependent Induction of a Dominant Negative Smad7 Splice Variant. PLoS One. 2016;11(8):e0161410.

281. Tao S, Sampath K. Alternative splicing of SMADs in differentiation and tissue homeostasis. Dev Growth Differ. 2010;52(4):335-42.

282. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2018;390(10114):2769-78.

283. Fang YJ, Lu ZH, Wang GQ, Pan ZZ, Zhou ZW, Yun JP, et al. Elevated expressions of MMP7, TROP2, and survivin are associated with survival, disease recurrence, and liver metastasis of colon cancer. Int J Colorectal Dis. 2009;24(8):875-84.

284. Ohmachi T, Tanaka F, Mimori K, Inoue H, Yanaga K, Mori M. Clinical significance of TROP2 expression in colorectal cancer. Clin Cancer Res. 2006;12(10):3057-63.

285. Stepan LP, Trueblood ES, Hale K, Babcook J, Borges L, Sutherland CL. Expression of Trop2 cell surface glycoprotein in normal and tumor tissues: potential implications as a cancer therapeutic target. J Histochem Cytochem. 2011;59(7):701-10.

286. Xu KY, Gu J. [Expression of TROP2 mRNA in left-sided and right-sided colon cancer and its clinical significance]. Zhonghua Wei Chang Wai Ke Za Zhi. 2009;12(3):285-9.

287. Wang J, Day R, Dong Y, Weintraub SJ, Michel L. Identification of Trop-2 as an oncogene and an attractive therapeutic target in colon cancers. Molecular Cancer Therapeutics. 2008;7(2):280.

288. Gandin V, Miluzio A, Barbieri AM, Beugnet A, Kiyokawa H, Marchisio PC, et al. Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation. Nature. 2008;455(7213):684-8.

289. Kesselring R, Glaesner J, Hiergeist A, Naschberger E, Neumann H, Brunner SM, et al. IRAK-M Expression in Tumor Cells Supports Colorectal Cancer Progression through Reduction of Antimicrobial Defense and Stabilization of STAT3. Cancer Cell. 2016;29(5):684-96.

290. Rothschild DE, Zhang Y, Diao N, Lee CK, Chen K, Caswell CC, et al. Enhanced Mucosal Defense and Reduced Tumor Burden in Mice with the Compromised Negative Regulator IRAK-M. EBioMedicine. 2017;15:36-47.

291. Foell D, Ichida F, Vogl T, Yu X, Chen R, Miyawaki T, et al. S100A12 (EN-RAGE) in monitoring Kawasaki disease. Lancet. 2003;361(9365):1270-2.

292. Foell D, Kane D, Bresnihan B, Vogl T, Nacken W, Sorg C, et al. Expression of the pro-inflammatory protein S100A12 (EN-RAGE) in rheumatoid and psoriatic arthritis. Rheumatology (Oxford). 2003;42(11):1383-9.

293. Foell D, Kucharzik T, Kraft M, Vogl T, Sorg C, Domschke W, et al. Neutrophil derived human S100A12 (EN-RAGE) is strongly expressed during chronic active inflammatory bowel disease. Gut. 2003;52(6):847-53.

294. Thierolf M, Hagmann ML, Pfeffer M, Berntenis N, Wild N, Roessler M, et al. Towards a comprehensive proteome of normal and malignant human colon tissue by 2-D-LC-ESI-MS and 2-DE proteomics and identification of S100A12 as potential cancer biomarker. Proteomics Clin Appl. 2008;2(1):11-22.

295. Zhang H, Wang Q, Zhao Q, Di W. MiR-124 inhibits the migration and invasion of ovarian cancer cells by targeting SphK1. J Ovarian Res. 2013;6(1):84.

296. Qin Z, Liu X. miR-124, a potential therapeutic target in colorectal cancer. Onco Targets Ther. 2019;12:749-51.

297. Liu K, Zhao H, Yao H, Lei S, Lei Z, Li T, et al. MicroRNA-124 regulates the proliferation of colorectal cancer cells by targeting iASPP. Biomed Res Int. 2013;2013:867537.

298. Oltra SS, Pena-Chilet M, Vidal-Tomas V, Flower K, Martinez MT, Alonso E, et al. Methylation deregulation of miRNA promoters identifies miR124-2 as a survival biomarker in Breast Cancer in very young women. Sci Rep. 2018;8(1):14373.

299. Zhou L, Xu Z, Ren X, Chen K, Xin S. MicroRNA-124 (MiR-124) Inhibits Cell Proliferation, Metastasis and Invasion in Colorectal Cancer by Downregulating Rho-Associated Protein Kinase 1(ROCK1). Cell Physiol Biochem. 2016;38(5):1785-95.

300. Bromberg J, Wang TC. Inflammation and cancer: IL-6 and STAT3 complete the link. Cancer Cell. 2009;15(2):79-80.

301. Hiltunen MO, Koistinaho J, Alhonen L, Myohanen S, Marin S, Kosma VM, et al. Hypermethylation of the WT1 and calcitonin gene promoter regions at chromosome 11p in human colorectal cancer. Br J Cancer. 1997;76(9):1124-30.

302. Magnani G, Furlan D, Sahnane N, Reggiani Bonetti L, Domati F, Pedroni M. Molecular Features and Methylation Status in Early Onset (</=40 Years) Colorectal

Cancer: A Population Based, Case-Control Study. Gastroenterol Res Pract. 2015;2015:132190.

303. Suzuki H, Romano-Spica V, Papas TS, Bhat NK. ETS1 suppresses tumorigenicity of human colon cancer cells. Proc Natl Acad Sci U S A. 1995;92(10):4442-6.

304. Sato F, Shibata D, Harpaz N, Xu Y, Yin J, Mori Y, et al. Aberrant methylation of the HPP1 gene in ulcerative colitis-associated colorectal carcinoma. Cancer Res. 2002;62(23):6820-2.

305. Klump B, Hsieh CJ, Holzmann K, Gregor M, Porschen R. Hypermethylation of the CDKN2/p16 promoter during neoplastic progression in Barrett's esophagus. Gastroenterology. 1998;115(6):1381-6.

306. Matkowskyj KA, Chen ZE, Rao MS, Yang GY. Dysplastic lesions in inflammatory bowel disease: molecular pathogenesis to morphology. Arch Pathol Lab Med. 2013;137(3):338-50.

307. Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res. 1998;58(23):5489-94.

308. Yang Q, Huang T, Ye G, Wang B, Zhang X. Methylation of SFRP2 gene as a promising noninvasive biomarker using feces in colorectal cancer diagnosis: a systematic meta-analysis. Sci Rep. 2016;6:33339.

309. Sun J, Zheng MY, Li YW, Zhang SW. Structure and function of Septin 9 and its role in human malignant tumors. World J Gastrointest Oncol. 2020;12(6):619-31.

310. Ma ZY, Law WL, Ng EKO, Chan CSY, Lau KS, Cheng YY, et al. Methylated Septin 9 and Carcinoembryonic Antigen for Serological Diagnosis and Monitoring of Patients with Colorectal Cancer After Surgery. Sci Rep. 2019;9(1):10326.

311. Yang X, Xu ZJ, Chen X, Zeng SS, Qian L, Wei J, et al. Clinical value of preoperative methylated septin 9 in Chinese colorectal cancer patients. World J Gastroenterol. 2019;25(17):2099-109.

312. Gorelik L, Flavell RA. Immune-mediated eradication of tumors through the blockade of transforming growth factor-beta signaling in T cells. Nature Medicine. 2001;7(10):1118-22.

313. Kaz AM, Wong CJ, Dzieciatkowski S, Luo Y, Schoen RE, Grady WM. Patterns of DNA methylation in the normal colon vary by anatomical location, gender, and age. Epigenetics. 2014;9(4):492-502.

314. Barnicle A, Seoighe C, Golden A, Greally JM, Egan LJ. Differential DNA methylation patterns of homeobox genes in proximal and distal colon epithelial cells. Physiol Genomics. 2016;48(4):257-73.

315. Luo Y, Wong CJ, Kaz AM, Dzieciatkowski S, Carter KT, Morris SM, et al. Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer. Gastroenterology. 2014;147(2):418-29 e8.