E-cadherin and APC in early-onset colorectal cancer

Alan Aitchison

A thesis submitted for the degree of

Doctor of Philosophy

University of Otago Christchurch

New Zealand

June 2020

Abstract

Colorectal cancer (CRC) is a considerable health burden being the second highest cause of cancer deaths globally. While overall cases of CRC have been declining worldwide, there has been an increase in the incidence of the disease among patients under 50 years of age. The majority of these cancers are sporadic and the increase in incidence may reflect changing lifestyle, exposing young people to more and earlier pro-oncogenic factors.

An early event in the development of CRC is the loss of normal structure of the epithelium and key to this is the loss of cell to cell contact. E-cadherin, encoded by the *CDH1* gene, is a membrane-bound protein whose extracellular domains bind to E-cadherin of neighbouring cells forming adherens junctions as the primary event in intercellular contact. Loss of E-cadherin leads to the breakdown of this organised epithelial structure and can lead to the development of cancer.

The research detailed in this thesis looked at E-cadherin expression status and mutation of the *CDH1* gene in two separate cohorts. Firstly tumours from young Pakistani patients with early-onset colorectal signet-ring cell carcinomas (SRCCs), due to their histological similarity to SRCCs found in hereditary diffuse gastric cancers where mutation of *CDH1* is a common causative factor and, secondly a local New Zealand cohort of early-onset CRC cases.

E-cadherin was found to be absent or weak in the colorectal SRCC samples and in a small number of cases this corresponded to germline *CDH1* mutation. However, the remaining SRCC samples had low levels of *CDH1* mutation despite loss of E-cadherin expression, suggesting that while E-cadherin loss is common in colorectal SRCC it is not due to *CDH1* mutation in most cases. Amongst the New Zealand samples, only one tumour, the sole SRCC case in the cohort, showed loss of E-cadherin but this was not correlated with *CDH1* mutation which was not a common feature of this cohort.

Additionally, *APC* sequencing was performed on the New Zealand cohort. *APC* is mutated in 60-80% of sporadic CRC tumours but with reportedly lower levels in younger patients. However 72% of the New Zealand early-onset CRC cohort was found to have a mutation in *APC*, a higher proportion than expected. This may reflect a greater coverage of *APC* by the sequencing methodology employed in this study compared to previous studies with a high proportion of mutations occurring outside the commonly studied mutation cluster region of

APC. Loss of heterozygosity at the *APC* locus was found in only three patients, all of whom had *APC* mutations occurring close to codon 1300, reflective of a previous studies in older-onset CRC.

While environmental and lifestyle factors are widely considered to have roles in the development of sporadic CRC, there is growing evidence of the gut microbiota being a factor in colorectal carcinogenesis. One well-studied toxin-producing bacteria is Enterotoxigenic *Bacteroides fragilis* (ETBF) that causes cleavage of E-cadherin in colonic epithelial cells. Study of the effect of the *B. fragilis* toxin (BFT) on colonic epithelial cells has focused on the cell line HT29 which shows a rapid morphological change upon incubation with BFT. However, HT29 cells only have truncated forms of APC. In the last part of this study one mutated *APC* allele in HT29 was corrected by genome editing in order to study the effects of BFT on a cell line expressing full-length APC, and thereby to increase our understanding of the role of APC in colorectal carcinogenesis.

The rapid change in morphology upon exposure to BFT in HT29 cells has been attributed to the BFT-mediated cleavage of E-cadherin. However, edited HT29 cells containing full-length APC maintained their structure after 6 hours incubation with ETBF supernatant. Moreover, fluorescent immunohistochemistry showed that cell morphology was maintained despite the cleavage of E-cadherin, suggesting that the structural integrity of the edited cells was due to some internal function of APC rather than E-cadherin cleavage.

In summary, this research found that E-cadherin loss was commonly found in colorectal SRCCs but predominantly occurred independently from *CDH1* mutations. Furthermore, other than in SRCCs, E-cadherin loss was not a common feature of early-onset CRC. Conversely, *APC* mutation was very common, with many mutations being found outside the mutation cluster region of *APC*, suggesting *APC* sequencing strategies should be targeted more widely throughout the gene. These results suggest a role for the APC protein in stabilising the cellular morphology of HT29 cells exposed to bacterial toxin, which is independent of E-cadherin. The edited HT29 cell line is likely to be a useful tool in the study of bacterial and other environmental effects on colorectal carcinogenesis.

Aknowledgements

I would like to express my immense gratitude to my primary supervisor, Jacqui Keenan, without whom this thesis would not have been written. Your unwavering support, patience and encouragement, not just through my PhD studies but throughout the last seven years I have been fortunate enough to work with you, has been invaluable and very much appreciated. I genuinely could not have wished for a better supervisor. My thanks also go to my co-supervisors Paul Pace, whose expertise made some of the technical aspects of this project navigable, and Martin Kennedy, whose knowledge and words of encouragement were greatly appreciated. I am in the enviable position of being able to regard all three supervisors as friends first and supervisors second.

Considerable thanks must also go to Chris Hakkaart for his help with next generation sequencing. His patient explanation of the intricacies and pitfalls of sequence analysis were gratefully appreciated. Also, to Helen Morrin for her help with supplying samples from the Cancer Society Tissue Bank and her accommodation of my requests despite her hectic schedule, I am extremely grateful.

I am indebted to wider support from all those within the Department of Surgery, particularly Frank Frizelle for his support, Rachel Purcell for her friendly encouragement and Morwenna Malcolm, Bailey Hamilton and Rebekah Higgs for their administrative support.

To those from the Departments of Surgery, Paediatrics, Medicine and Haematology who share our laboratory, past and present, I thank you. To Judy, Barry, Shaun, Phoebe, Teagan, Liping, Will and all the others it has been a pleasure to have had your encouragement during these years as well as the distracting but hilarious conversations. This journey has been long and taken many unexpected turns, both scientifically and personally. To have a great bunch of colleagues to keep me grounded has been invaluable. To Shaun, Phoebe and Will, I look forward to meeting you on the other side in the near future.

Finally, to Cameron, Ewan and Finlay, I could not be prouder of the way you have developed over the past few years. You have all dealt with whatever life has thrown at you with humility, strength and a huge dose of good humour. Thank you for keeping me grounded. I owe you a huge debt which I now intend to repay in spades.

Table of Contents

List of Figures

List of Tables

Abbreviations

APC	Adenomatous polyposis coli
BFT	*Bacteroides fragilis* toxin
BSA	Bovine serum albumin
CRC	Colorectal cancer
CRISPR	Clustered regularly interspaced palindromic repeats
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraaecetic acid
EMT	Epithelial-mesenchymal transition
ETBF	Enterotoxigenic *Bacteroides fragilis*
FBS	Fetal bovine serum
FFPE	Formalin-fixed paraffin-embedded
gRNA	Guide RNA
IF	Immunofluorescence
IHC	Immunohistochemistry
LOH	Loss of heterozygosity
MMR	Mismatch repair
MPW	Millipore water
NTBF	Non-enterotoxigenic *Bacteroides fragilis*
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
PVDF	Polyvinylidene difluoride
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
SRCC	Signet-ring cell carcinoma

TBE	Tris borate EDTA
TE	Tris-EDTA solution
v/v	Volume to volume dilution
w/v	Weight to volume dilution
WB	Western Blotting
WHO	World Health Organisation

1 Introduction

1.1 Colorectal Cancer

Colorectal cancer (CRC) is a considerable health burden globally, being the second most diagnosed cancer in women and third in men (Ferlay et al., 2015). New Zealand ranks amongst the highest in the world for incidence (Gandhi et al., 2017) and CRC is the second highest cause of cancer-related death (Gandhi et al., 2017).

CRC is influenced by both genetic and environmental factors. While hereditary syndromes predisposing patients to CRC, for example hereditary non-polyposis colorectal cancer (Lynch et al., 1991) and familial adenomatous polyposis (Groden et al., 1991; Nishisho et al., 1991) exist, the majority of CRC cases are sporadic.

1.1.1 Hereditary colorectal cancer syndromes

1.1.1.1 Hereditary Non-Polyposis Colorectal Cancer

Hereditary non-polyposis colorectal cancer (HNPCC), also known as Lynch syndrome, is an autosomal dominant condition caused by mutations in DNA mismatch repair genes (Lynch et al., 1991). DNA damage can occur via errors in DNA replication or by the effects of exogenous factors, such as chemicals, cigarette smoke and radiation, or endogenous metabolites including reactive oxygen or nitrogen species (G. M. Li, 2008). To prevent such damage affecting the integrity of the genome, cells possess multiple mechanisms to repair this damage, an important one of which is DNA mismatch repair (MMR). Loss of activity of MMR proteins leads to rapid accumulation of DNA replication errors, including within key regulatory genes involved in cell control, apoptosis and DNA repair, increasing the potential for tumour formation (Boland et al., 2008). The four principal MMR genes mutated in HNPCC are MutL homolog 1 (MLH1), post-meiotic segregation 2 (PMS2), MutS homolog 6 (MSH6) and MutS homolog 2 (MSH2) (Al-Sohaily et al., 2012). Mutation of a further gene that encodes an epithelial cell adhesion molecule Ep-CAM (TACSTD1) has also been implicated in HNPCC due to its silencing effect on MSH2 (Ligtenberg et al., 2009).

The accumulation of DNA replication errors in HNPCC is particularly apparent in microsatellites, short regions of repetitive sequences that can acquire losses or gains of the

repeated sequence. This manifests itself as microsatellite instability (MSI) and microsatellite testing can be performed by amplifying these short repetitive regions and looking for changes in the lengths of the amplicons due to the loss or gain of repeat sequences (Suraweera et al., 2002). Alternatively, HNPCC can be diagnosed using immunohistochemistry by measuring the loss of expression of any of MSH2, MSH6, MLH1 or PMS2 (South et al., 2009). This will also be effective in the case of mutation of TACSTD1 as this also results in the loss of MSH2 expression.

HNPCC is the most common inherited form of CRC, accounting for 2-3% of all colorectal cancer cases (Desai & Barkel, 2008). Tumours occur at an earlier age in HNPCC patients than in the general population (Cunningham et al., 2010), and tend to be located on the right side of the colon and often occur with synchronous and metachronous colorectal tumours (Vasen, 2007).

1.1.1.2 Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP) accounts for less than 1% of all CRC cases and, despite being less common than HNPCC, was the first familial colorectal cancer syndrome identified due to its striking clinical characteristics. The disease is typified by the early-onset of hundreds to thousands of adenomatous polyps within the colon, and if left untreated almost 100% of patients will develop colorectal cancer by the age of 40 years (Bisgaard et al., 1994). Cancer can be prevented by rigorous screening and surgical intervention. FAP is an autosomal dominant disease caused by mutations in the adenomatous polyposis coli (*APC*) gene (Groden et al., 1991). A detailed analysis of the role of *APC* in CRC is introduced below.

Attenuated FAP is a less aggressive form of the disease with a later age of onset and fewer adenomatous polyps. While attenuated FAP is also caused by mutations in *APC*, the causal mutations appear confined to three regions of *APC*: the 5' region comprising the first 5 exons, exon 9 and the 3' region of the gene (Knudsen et al., 2003). It is estimated around 8% of FAP cases may present with an attenuated FAP phenotype (Nielsen et al., 2007).

1.1.1.3 Other inherited colorectal cancer syndromes

A number of other conditions can result in rare inherited forms of CRC. *MUTYH*-associated polyposis is an autosomal recessive sub-type that occurs as a result of mutation of the *MUTYH* gene, which encodes the MYH glycosylase, which is involved in DNA base excision repair. Loss of MYH glycosylase leads to the accumulation of DNA mutations in the cell and subsequent tumour formation (Poulsen & Bisgaard, 2008).

Peutz-Jeghers syndrome is an autosomal dominant condition caused by mutation of the serine/threonine kinase 11 gene (*STK11*, also known as *LKB1*) (Jenne et al., 1998) with an incidence of between 1 in 25,000 and 1 in 300,000 people. *STK11* is a tumour suppressor gene that regulates cell polarity and promotes programmed cell death. Mutation of *STK11* is associated with uncontrolled epithelial cell growth, leading to the formation of hamartomatous polyps and a 41-60% risk of malignant tumours (Vaahtomeri & Makela, 2011).

Cowden syndrome, an inherited condition affecting approximately 1 in 200,000 people, is caused by mutation of the phosphatase and tensin homolog gene (*PTEN*), and manifests as an increased risk of multiple malignancies including colon cancer. Multiple hamartomatous polyps are a common feature in Cowden syndrome (Stanich et al., 2011). Interestingly the *STK11* and *PTEN* genes implicated in the formation of hamartomatous polyposis syndromes in Peutz-Jeghers and Cowden syndromes, respectively, have been shown to interact with each other, suggesting these two syndromes may develop through similar pathways (Mehenni et al., 2005).

1.1.2 Sporadic colorectal cancer pathways

Approximately 75% of colorectal cancer is sporadic, occurring in patients without a genetic predisposition or family history of CRC. Tumour formation is the result of an accumulation of genetic and epigenetic alterations (Yamagishi et al., 2016). Over the past few decades three pathways of carcinogenesis, comprising different but overlapping patterns of mutated or altered genes, have been proposed (Figure 1.1).

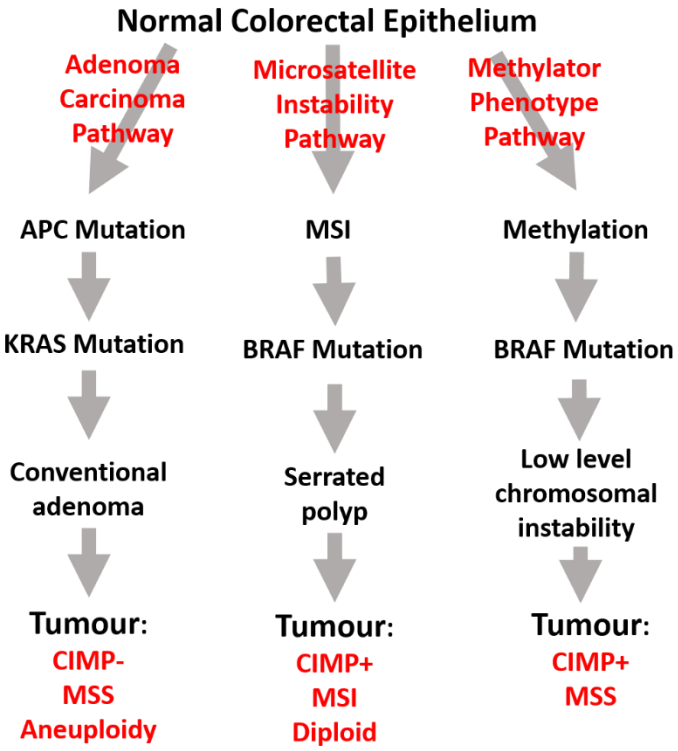

Figure 1.1 Sporadic colorectal cancer pathways. Diagram showing the key genetic and epigenetic events in each of the three pathways of colorectal cancer development. APC, adenomatous polyposis coli gene; KRAS, Kirsten rat sarcoma proto-oncogene; BRAF, v-RAF murine sarcoma viral oncogene homolog B; MSI, microsatellite instability; MSS microsatellite stable; CIMP, CpG island methylator phenotype.

1.1.2.1 The adenoma-carcinoma pathway

The genes mutated in familial CRCs are also key to sporadic CRC, with mutation of *APC* in particular being an early event in colorectal carcinogenesis. *APC* mutations are found in the earliest microscopic adenomas (Kinzler & Vogelstein, 1996) and *APC* mutations in sporadic tumours are found at the same rate as in early adenomas (Polakis, 2007). Additionally, both alleles of *APC* appear to require inactivation for carcinogenesis to occur, thereby fulfilling Knudsen's two-hit hypothesis (Kinzler & Vogelstein, 1996). However, *APC* mutation alone is not sufficient in itself for CRC progression as not all adenomas will progress to carcinomas, as was evidenced by a study that found no significant difference in *APC* status between progressed and non-progressed adenomas (Hermsen et al., 2002). Those adenomas that do progress to CRCs acquire an accumulation of additional genetic aberrations such as mutations of the Kirsten rat sarcoma proto-oncogene (*KRAS*), V-RAF murine sarcoma oncogene homolog B1

(*BRAF*), SMA- and MAD-related proteins 2 and 4 (*SMAD2/4*), and tumour protein 53 (*TP53*), that occur over time as dysregulation continues (Sameer, 2013). These are often referred to as the adenoma-carcinoma sequence (Figure 1.2; (Fearon & Vogelstein, 1990)).

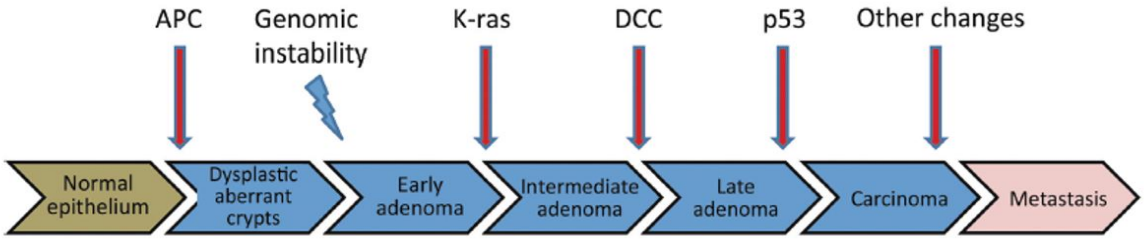

Figure 1.2 The adenoma-carcinoma model of colorectal carcinogenesis. The sequential accumulation of genetic aberrations proposed by Fearon and Vogelstein starting with mutation of the APC gene. *DCC*, Deleted in colon cancer (Figure taken from Yalcin, 2014).

An early event following *APC* mutation in the adenoma-carcinoma pathway is genetic disruption in the form of loss or gain of whole arms, or significant portions of arms of chromosomes in what is termed chromosomal instability (CIN) (Pino & Chung, 2010). This results in an imbalance in chromosome number (aneuploidy) and a high frequency of loss of heterozygosity (LOH). Common chromosomal anomalies in CIN include loss of chromosome 18, partial loss of chromosomes 1, 5, 8 and 17, as well as gains and losses of large parts of various chromosomes (Sheffer et al., 2009; Thiagalingam et al., 2001). Of note, loss of chromosome 5q, the chromosome arm carrying the *APC* gene, is seen in 20-50% of colorectal cancers (Fearon & Vogelstein, 1990).

1.1.2.2 The microsatellite instability pathway

While mutation of the *APC* gene is a hallmark of the adenoma-carcinoma pathway, *APC* mutation *per se* is not an absolute requirement for CRC progression as alternative pathways of CRC exist, as evidenced by the finding that a minority of CRC tumours present with wild type *APC*. This group includes the 15 to 30% of CRC cases that develop from serrated polyps, characterised by their saw-tooth appearance (Bettington et al., 2013). Neoplastic lesions arising from serrated polyps rarely present with *APC* mutations but commonly have mutations in

BRAF and to a lesser extent *KRAS* (De Palma et al., 2019). Another common feature of the serrated pathway is microsatellite instability (De Palma et al., 2019).

Microsatellite instability (MSI), the result of mutations in mismatch repair genes, manifests as the destabilisation of repetitive tracts of DNA and is the underlying cause of hereditary non-polyposis colorectal cancer, as detailed above. MSI is also seen in approximately 15% of sporadic colorectal cancers (Boland & Goel, 2010). As with familial HNPCC, mutation of mismatch repair genes leads to the rapid accumulation of DNA errors, including aberrations in key genes that regulate the cell cycle, apoptosis and DNA repair (Boland et al., 2008).

The presence of MSI in CRCs is associated with an improved prognosis compared to microsatellite stable (MSS) tumours (Popat et al., 2005). The reasons for this are unclear. Notably, tumours with MSI tend to have less loss of heterozygosity than tumours in the chromosomal instability pathway (Soreide et al., 2006). Furthermore, mutations or allele losses in *KRAS*, *TP53* and *DCC* (Deleted in Colon Cancer) are associated with a poorer prognosis and mutations within these genes are uncommon in MSI tumours (Soreide et al., 2006). Conversely, BRAF V600E mutations are common in MSI tumours while they are not seen in HNPCC tumours (G. Deng et al., 2004). In addition, MSI tumours are observed to be associated with lymphocyte infiltration, and it is possible that this immune response may contribute to a better outcome (Linnebacher et al., 2010).

1.1.2.3 The methylator phenotype pathway

The discovery during the 1990's, that multiple suppressor genes are silenced in colorectal tumours, led to the development that epigenetic aberrations can define tumour progression as a CpG island methylator phenotype (CIMP) (Toyota et al., 1999), where expression of key tumour suppressor genes are repressed by promoter methylation (Goel et al., 2007). This led to the assertion that CIMP could be used as a classifier and prognostic indicator in CRCs. While the majority of colorectal tumours have genetic aberrations due to either MSI or CIN, the CIMP can occur alongside CIN and is often associated with MSI, but appears to arise from an independent mechanism of tumour progression (Cheng et al., 2008).

Tumours exhibiting a CIMP account for about 15-20% of sporadic CRCs, are more common in women and older patients, and are often located in the proximal colon (Nosho et al., 2008).

CIMP-positive tumours are often associated with MSI and *BRAF* mutation (Weisenberger et al., 2006) and this may be due to methylation of the *MLH1* promoter.

Environmental factors such as smoking (Samowitz et al., 2006) and aging (Toyota & Issa, 1999) have been shown to correlate with increased methylation. A detailed discussion of environmental and lifestyle factors, and their effect on epigenetics and carcinogenesis, is presented below (Section 1.7).

1.2 The structure of the colonic epithelium

The colorectal epithelium consists of a monolayer of polarised cells that act as a barrier, keeping the rich microbe community of the gut lumen from infecting the rest of the body. The formation of this barrier starts at the base of colonic crypts with continual division of a small population of stem cells. At each stem cell division, one cell becomes a transit amplifying cell that continues to proliferate, populating first the crypt and then starting to migrate up the crypt towards the villus (Boman & Fields, 2013). As numbers increase, cells migrate up the crypt where there is a reduction in proliferation as the cells polarise and then differentiate (Pinto & Clevers, 2005).

Central to the establishment of this cell polarity is the formation of adherens junctions (Knust & Bossinger, 2002) between adjacent cells where the intracellular protein E-cadherin forms homocomplexes with its neighbouring cell (Figure 1.3). The binding of E-cadherin not only contacts neighbouring cells but also provides a structural link with intracellular structures via catenin proteins (Paredes et al., 2012). The resultant polarised cell monolayer moving up the crypt has clearly defined basolateral and apical membranes (Wodarz & Nathke, 2007). Tight junctions then form between adjacent cells and integrins bind the basal membrane to the basement membrane. The formation of this E-cadherin-mediated barrier not only provides structure to the cells, but also acts as a signal to suppress further cell proliferation (Mendonsa et al., 2018).

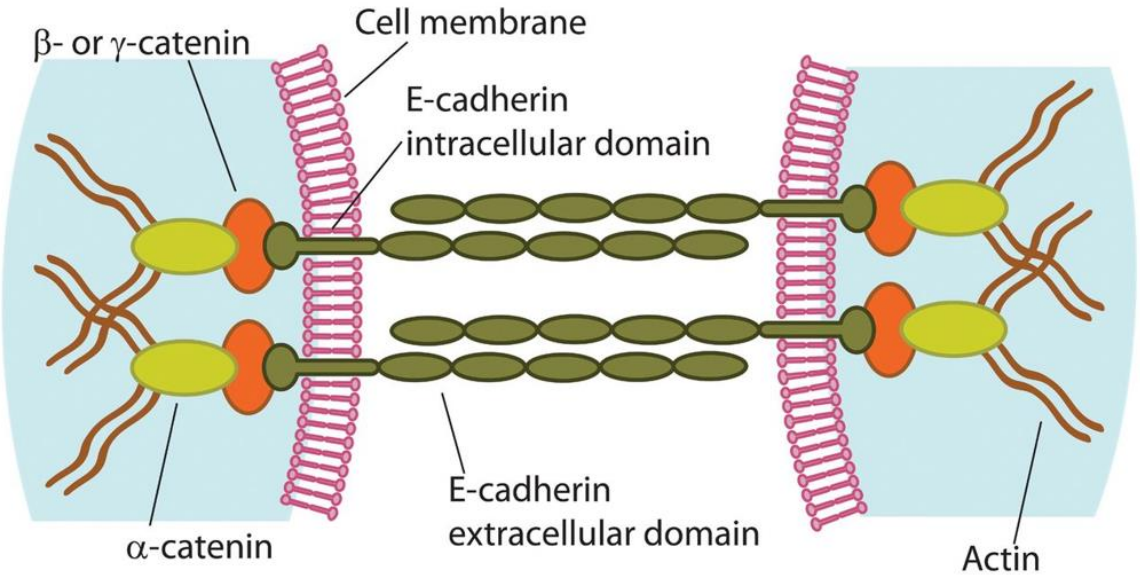

Figure 1.3 Binding of the extracellular domains of E-cadherin proteins from adjacent cells. E-cadherin binds adjacent cells at adherens junctions while connecting intracellularly to the actin cytoskeleton via catenin molecules (Figure taken from Perry et al., 2009).

The polarised cells continue to migrate up the crypt, gradually differentiating until they are shed from the surface of the villus into the lumen of the gut, after a crypt to villous migration of approximately five days (Potten et al., 1992).

1.3 The cytoskeleton

The intracellular domain of E-cadherin binds to β/γ-catenin in a complex with α-catenin which attaches to the actin cytoskeleton (Figure 1.3) (Takeichi, 1991). The cytoskeleton is a dynamic fibrous network throughout the cytoplasm that regulates cellular architecture as well as providing a scaffold for diverse biochemical pathways (S. Kim & Coulombe, 2010). Structurally the cytoskeleton is comprised of actin filaments and microtubules made from polymerised tubulin. As the cytoskeleton forms a dynamic network across the entire cell it is in contact with many cellular components such as the nucleus, various organelles, vesicles and proteins, many of which depend on the cytoskeleton for their function. Accordingly, the cytoskeleton is a regulator of many cellular processes including cell division, organelle positioning, vesicular trafficking, cell migration and adhesion (M. C. Kim et al., 2012).

The microtubule component of the cytoskeleton is stabilised by APC binding to its end with APC conferring a longer lifespan on microtubules (Kita et al., 2006). APC further strengthens the cytoskeleton by facilitating crosstalk between microtubules and actin filaments as a part of the cortical microtubule stabilisation complexes (Dogterom & Koenderink, 2019). The microtubule binding domain of APC is located near the C-terminal end of the protein which is usually lost in the truncated cancer-associated isoforms of APC. Thus disruption of the cytoskeletal stabilising role of APC may be one of the key mechanisms by which carcinogenesis is promoted by truncated APC (Nathke, 2006), in addition to its role in the regulation of Wnt signalling (Section 1.4.1).

1.4 Wnt signalling

Wnt (Wingless-related integration site) signalling is the dominant force driving epithelial cell proliferation in the colon. Wnt ligands bind to the transmembrane receptors Frizzled and LRP5/6 (X. He et al., 2004). Binding of the ligands stabilises β-catenin intracellularly which then moves to the nucleus where it binds to T-cell factor (TCF) family transcription factors. The DNA-binding specificity of TCFs and the transactivation domains of β-catenin allow transcription of Wnt target genes such as the *MYC* oncogene leading to cell proliferation (Pinto & Clevers, 2005).

The effect of Wnt signalling on cell proliferation is desirable in the cells at the bottom of the colonic crypts where proliferation is key to the renewal of epithelial cells and cell turnover. However, as the cells move up the crypt and mature, Wnt-controlled cellular proliferation is less desirable and, accordingly, Wnt signalling becomes gradually inactive towards the top of the crypt in the normal epithelium (Daulagala et al., 2019).

The key intracellular regulator of Wnt signalling is the tumour suppressor protein adenomatous polyposis coli (APC). Consequently, the loss of functional APC, as seen in FAP (Groden et al., 1991) or in 70-80% of sporadic CRC (Fearon, 2011), is associated with dysregulated Wnt signalling (Figure 1.4).

1.4.1 APC function

APC is a multi-functional protein in epithelial cells, having roles in Wnt signalling (Boman & Fields, 2013), cell structure through interactions with the actin cytoskeleton and microtubules (Munemitsu et al., 1994), epithelial cell polarity (Bellis et al., 2012) and apoptosis (Cristofaro et al., 2015). As such APC has a number of different binding partners. APC is an important regulator of Wnt signalling. In the absence of a Wnt ligand, APC forms a complex with axin and GSK3β, recruiting β-catenin to the complex for its phosphorylation by GSK3β, resulting in the ubiquitination and, ultimately, proteasomal degradation of β-catenin. This prevents excess β-catenin translocating to the nucleus and upregulating oncogenic target genes such as *MYC* (Figure 1.4).

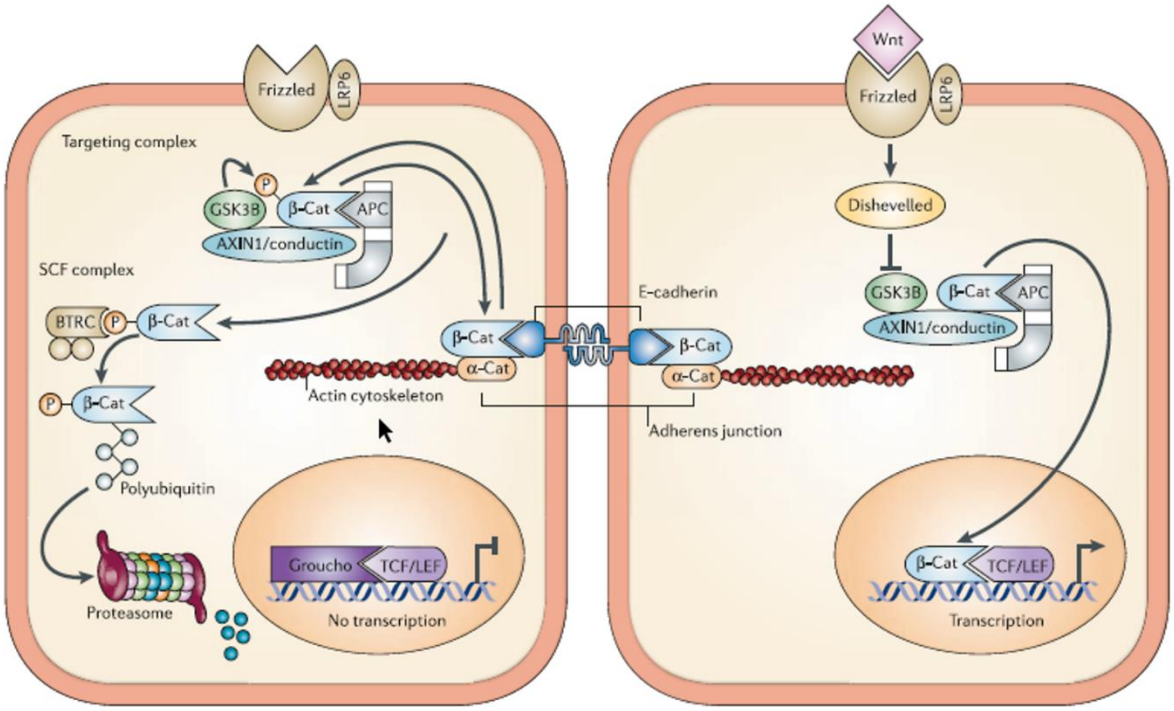

Figure 1.4 The Wnt signalling pathway. In the absence of Wnt ligand (left plate) β-catenin is phosphorylated by the APC complex and targeted for degradation. When Wnt ligand binds to its receptor (right plate), the phosphorylation of β-catenin is prevented, allowing nuclear translocation of β-catenin, its binding to TCF factor, and the transcription of target genes. This figure also highlights the interplay between the dual roles of β-catenin in Wnt signalling and intercellular binding via E-cadherin (Figure taken from Nathke, 2006).

The consequent effects of unrestricted upregulation of oncogenic Wnt target genes, caused by defects in APC, are exemplified in both hereditary familial adenomatous polyposis (FAP) and in sporadic CRCs (Miyoshi et al., 1992), which commonly carry mutations in the *APC* gene.

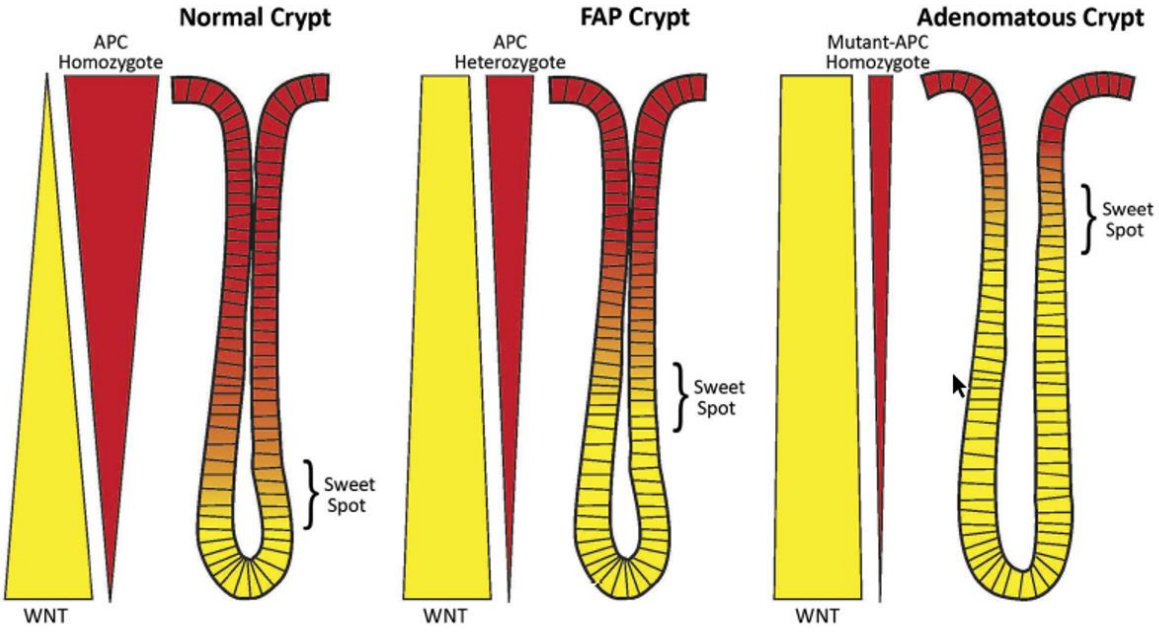

Figure 1.5 Wnt and APC gradients in colonic crypts. APC and Wnt are both essential for the regulation of cell proliferation and epithelial homeostasis in the gut. Wnt signalling is highest in the cells at the bottom of the crypt while APC levels increase as the cells move up the crypt. A sweet spot exists where conditions are optimal for balanced growth and cell proliferation. In the case of mutated or lost *APC*, with either one allele (FAP crypt) or both (adenomatous crypt), the sweet spot is higher in the crypt leading to uncontrolled proliferation of crypt cells (Figure taken from Boman & Fields, 2013).

The *APC* gene is located on chromosome 5q21-q22. Although it consists of 15 exons the final exon makes up more than 75% of the coding sequence of the gene. This is where the bulk of mutations occur, with a mutation cluster region between codons 1286 and 1513 (Miyoshi et al., 1992). Mutations in *APC* predominantly result in truncated proteins that retain some function and it appears that the retained functions are essential for the survival of the cells (Chandra et al., 2012). The second allele is either mutated as well or is silenced by loss of heterozygosity (Lamlum et al., 1999) or epigenetic mechanisms (Esteller et al., 2000).

The location of the truncating mutation affects the properties of the resultant protein and the tumour phenotype (Christie et al., 2013). While familial *APC* mutations are scattered across the 5' half of the gene, with notable hotspots at codons 1061 and 1309, sporadic mutations tend

to occur within the mutation cluster region (MCR). This region contains a number of 20-amino acid repeats that act as β-catenin binding sites. The location of a mutation within the MCR determines how many of these repeats are included in the truncated protein, with either 1, 2 or 3 intact 20-amino acid repeats being left. This in turn affects the function of the truncated protein, as well as the type and location of the second hit to the *APC* gene (Christie et al., 2013). This is exemplified by evidence that *APC* mutations in the MCR that result in 2-3 intact 20-amino acid repeats are more likely to be found in proximal tumours, while mutations leaving only 1 or no 20-amino acid repeats tend to be present in distal CRC (Christie et al., 2013). However, while the majority of sporadic *APC* mutations are found in the MCR, other hotspots have been found in the 5' portion of the gene, including codons 213, 216, 232, 283, 876 and 935, along with a splice-site mutation at c.835-8A>G (Christie et al., 2013).

1.4.2 E-cadherin

The development of dysplasia from normal epithelium is key to tumour formation in the gut (Tjalsma et al., 2012). An important early step is the loss of cell polarity of the epithelium. Loss of E-cadherin at the plasma membrane results in loss of intercellular contact at adherens junctions, and ultimately loss of cell polarity. This breakdown in the organised structure of the epithelium allows dysplastic growth of epithelial cells which can lead to the formation of adenomas (Figure 1.6).

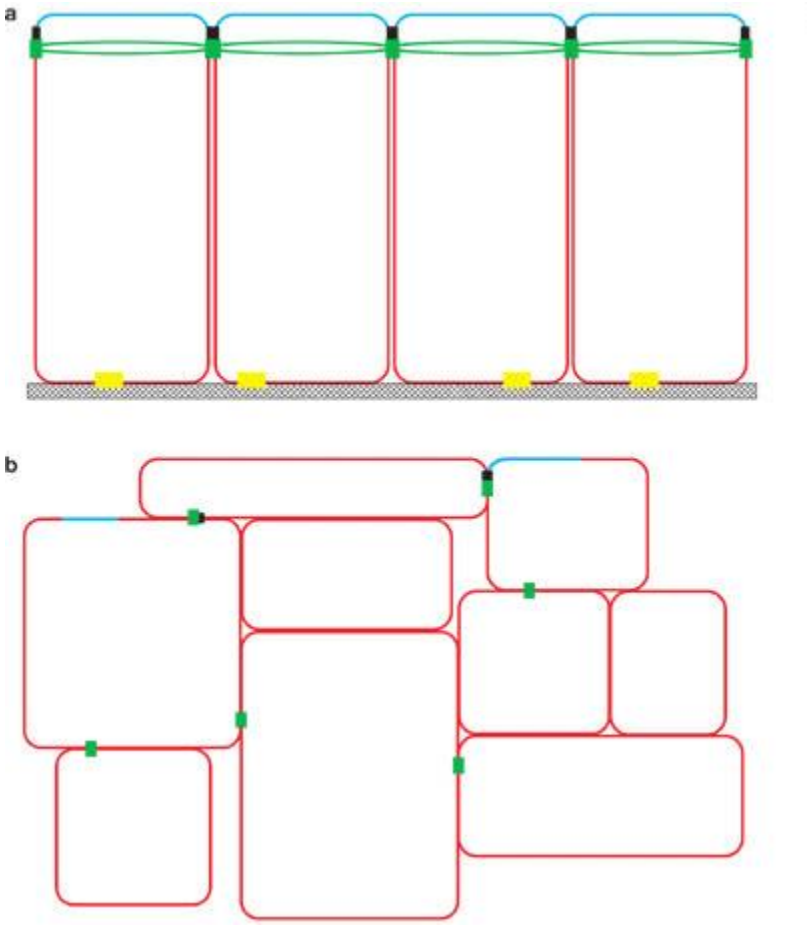

Figure 1.6 Loss of E-cadherin binding results in changes to cellular architecture. A, The normal epithelium consists of a monolayer of polar cells joined to neighbouring cells by adherens (green) and tight junctions (black) on the lateral membrane. The binding of E-cadherin forms the adherens junctions in a ring around the cell (green lines). The apical layer (blue) faces the lumen. B, Loss of E-cadherin causes disintegration of the adherens junctions, reduced cell-cell adhesion, loss of cell polarity and results in heterogenous cell sizes and shapes (Figure taken from Wodarz & Nathke, 2007).

A further consequence of the loss of cell polarity due to a loss of E-cadherin binding is enhanced Wnt signalling. This results in the transcription of pro-oncogenic Wnt target genes, in particular *MYC,* which encodes the transcription factor c-myc, whose upregulation is linked to stimulation of polyamine metabolism via upregulation of spermine oxidase (SMO) expression. This in turn leads to DNA damage via the production of reactive oxygen species (Goodwin et al., 2011).

E-cadherin, encoded by the *CDH1* gene, is a calcium-dependent mediator of cell to cell adhesion, facilitating the assembly of intercellular junctions (Gumbiner et al., 1988; Shiozaki et al., 1996), an essential step in forming an epithelial cell monolayer and induction of cell

polarity (Yap et al., 1995) as described above. Accordingly a loss of E-cadherin activity disrupts cell layers (Takeichi, 1990).

Mutations in the *CDH1* gene are rare in CRC but are commonly seen in hereditary diffuse gastric cancer (HDGC) (Guilford et al., 1998; Hakkaart et al., 2019). HDGC tumours are typified by the presence of diffuse foci of signet-ring cells (Charlton et al., 2004), where a large amount of intracellular mucin distorts the cell morphology, pushing the nucleus to the edge of the cell giving the classic signet-ring appearance to the cell. Loss of E-cadherin disrupts intercellular structure allowing the change in morphology seen in these tumours. Accordingly, signet-ring cells show a loss of E-cadherin protein expression (H. C. Kim et al., 2002).

Loss of E-cadherin has been shown in colorectal cancer, although this is not always as a result of reduced expression but rather relocation from the membrane to the cytoplasm (Hiscox & Jiang, 1997). In a recent study lower expression of E-cadherin was found to be associated with tumour differentiation, stage, invasion depth and lymph node metastasis in advanced CRC patients (Gao et al., 2016) and accordingly E-cadherin has been proposed as a possible biomarker of CRC prognosis (Christou et al., 2017).

1.5 Risk factors for colorectal cancer

1.5.1 Gut bacteria

The human gastrointestinal tract is home to tens of trillions of microbes, with each human carrying in excess of 1000 different species (Lozupone et al., 2012). This diverse population is acquired by babies from their mothers at birth (Dominguez-Bello et al., 2010) and from the environment where they spend the first few years of their life before stabilising (Kostic et al., 2013). Each individual's unique microbiome is essential in maintaining a healthy gut, with individual species complementing the host cells by digestion of complex carbohydrates (Flint et al., 2012), producing essential nutrients such as vitamin K (Conly & Stein, 1992), maintaining complex immunological interactions with host cells (Littman & Pamer, 2011), and protecting against disease by outcompeting pathogens (Wardwell et al., 2011). However, some commensal bacteria can, under the correct conditions, initiate disease.

The oncogenic potential of bacteria has been known since the discovery of *Helicobacter pylori* as a cause of gastric cancer in the 1980s (Marshall et al., 1985). However, whereas *H. pylori*

inhabits the microbially sparse environment of the stomach and acts alone in carcinogenesis, the crowded, diverse microbial community of the colon makes associations and mechanisms of bacterial carcinogenesis harder to ascertain.

The development of dysplasia is influenced by the inflammatory state of the gut, as clearly seen in Inflammatory Bowel Disease (IBD)-associated CRC (Beaugerie & Itzkowitz, 2015). Thus, sub-clinical inflammation may, over time, contribute to the development of dysplasia in an otherwise healthy colon. Moreover, long-term carriage of bacterial species that have the potential to cause chronic inflammation are increasingly considered to underlie the development of neoplasia (Armstrong et al., 2018).

One model of microbial initiation of colorectal cancer proposes distinct roles for different bacterial species. Initially "driver" bacteria cause inflammation, cell proliferation and/or the production of genotoxins that cause mutations in key tumour suppressor genes involved in the adenoma-carcinoma pathway such as *APC* and, later, *TP53* (Figure 1.7). The proliferation of cells and the breakdown of normal epithelial structure leads to changes in the local microenvironment which facilitate the replacement of these "driver" bacteria with other opportunistic "passenger" species with competitive advantages in the developing tumour (Tjalsma et al., 2012).

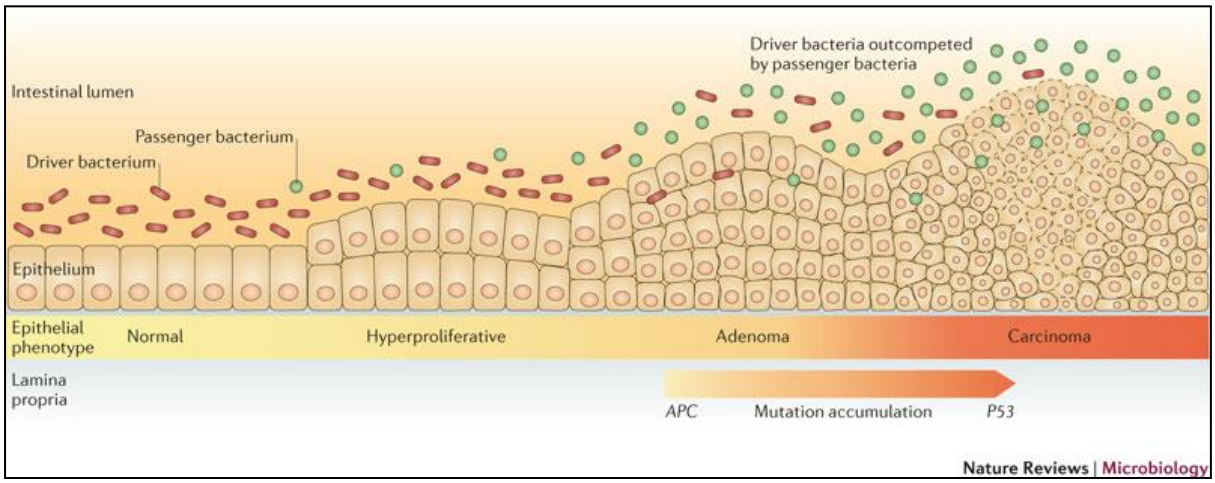

Figure 1.7 The driver-passenger model of bacterial carcinogenesis. In this model a driver bacterium disrupts the epithelial cell monolayer and encourages dysplasia, creating the conditions for a passenger bacterium that ultimately outcompetes the initial driver bacterium as the tumour microenvironment evolves (Figure taken from Tjalsma et al., 2012).

An alternative model postulates that certain bacterial species with unique virulence traits are both directly pro-oncogenic and capable of remodelling the microenvironment and local microbial community to further induce pro-oncogenic immune responses that also favour its own survival. As these bacteria do not work alone but rather co-opt other species this model has been termed the alpha-bug model (Sears & Pardoll, 2011).

A number of bacterial species have been associated with CRC by being found more commonly in the mucosa or stools of CRC patients than those of healthy controls, such as *Fusobacterium nucleatum* (Viljoen et al., 2015), colibactin-producing *Escherichia coli* (Buc et al., 2013) and enterotoxigenic *Bacteroides fragilis* (J. I. Keenan et al., 2016; Toprak et al., 2006). *Fusobacterium nucleatum* binds to colonic epithelial cells via FadA, an adhesin which recognises an 11-amino acid region of the extracellular domain of E-cadherin. Binding of FadA to E-cadherin upregulates β-catenin signalling leading to upregulation of transcription factors, oncogenes, Wnt genes and inflammatory genes (Rubinstein et al., 2013). The toxin produced by enterotoxigenic strains of *B. fragilis* is also associated with perturbation of E-cadherin and, while the overall effect is similar, the mechanism relates instead to toxin-mediated perturbation of intracellular signalling pathways (Wu et al., 2003). In contrast, strains of *E.coli* that contain a polyketide synthesis (pks) island are shown to cause enterocyte DNA damage in infected mice (Cuevas-Ramos et al., 2010). The pks island hosts a number of genes that encode the colibactin, a genotoxin that has been shown to promote tumour cell proliferation (Cougnoux et al., 2014). Interestingly, the FadA adhesin appears to be constitutively expressed by all *F. nucleatum* strains (Han et al., 2005). In contrast, only a subset of *B. fragilis* and *E. coli* strains express enterotoxin and colibactin respectively.

Whether these bacteria, detected in patients with advanced cancer, are important in the initiation or progression of cancer, or present as a consequence of favourable conditions in the tumour environment, has been a matter of considerable discussion. For example, multiple studies have found an association of *F. nucleatum* with CRC based on an increased presence within colorectal tumours (Castellarin et al., 2012; Kostic et al., 2012; McCoy et al., 2013). Our research however finds that colonic carriage of *F. nucleatum* is not associated with the subsequent growth of neoplastic lesions (Aitchison, unpublished), suggesting that *F. nucleatum* is more likely to be a passenger rather than a driver of colorectal carcinogenesis (Tjalsma et al., 2012).

Conversely, in the same patient cohort, our group has shown that long-term carriage of enterotoxigenic *B. fragilis* (ETBF) predisposes patients to early-stage colorectal neoplasia, providing evidence of ETBF as an initiator of carcinogenesis (Purcell et al., 2017). Accordingly, understanding how these toxin-producing colonic bacteria have the potential to drive this process is integral to our ongoing research.

1.5.1.1 Enterotoxigenic *Bacteroides fragilis*

Bacteroides fragilis are obligate anaerobes that are commonly found in the healthy human colon where they are particularly good at adhering to the mucosa (Sears et al., 2014), constituting about 1-2% of the normal microflora (Moore et al., 1978). However, enterotoxigenic strains of *B. fragilis* (ETBF) are associated with human and animal disease. ETBF were initially identified in newborn lambs with diarrhoea (Myers et al., 1984) and subsequently identified in multiple animal species with diarrhoeal disease, including a report of a piglet with exfoliating colitis with severe neutrophilic mucosal infiltrate (Collins et al., 1989), the first association of ETBF with inflammatory disease. In humans, ETBF was first isolated from eight patients with diarrhoea (Myers et al., 1987), while a later study of ETBF as a cause of diarrhoeal disease in children found that ETBF isolated from family members were genetically similar. Moreover, samples from the same patients taken months apart contained genetically similar ETBF, whether from diarrhoeal or normal stools, suggesting sustained infection (Sack et al., 1992).

The ETBF virulence factor is a zinc metalloprotease enterotoxin, commonly referred to as the *Bacteroides fragilis* toxin (BFT). Metalloproteases are enzymes commonly produced by bacteria and BFT is structurally similar to other bacterial toxins important in human disease such as tetanus toxin and diphtheria toxin (Sears & Pardoll, 2011).

BFT binds to a specific colonic epithelial cell membrane receptor, inducing γ-secretase-dependent cleavage of E-cadherin (Wu et al., 2007). As detailed above, E-cadherin forms complexes with α-, β-, and γ-cadherins at the cell membrane. Consequently, BFT-mediated cleavage of E-cadherin releases β-catenin to translocate to the nucleus where it helps transcribe multiple tumour-promoting genes as a key component of the Wnt signalling pathway (Wu et al., 2003).

In vitro BFT-mediated E-cadherin cleavage causes increased permeability of colonic mucosa whereby the colonic submucosa is exposed to normally luminal bacteria (Riegler et al., 1999), while *in vivo* this results in disruption of the colonic epithelium, followed by a subsequent inflammatory response and colitis in mice (Rhee et al., 2009). Stimulation of immune cells by bacteria or bacterial antigens may create an inflammatory environment that leads to carcinogenesis. BFT affects the immune system in a number of ways, including regulatory T-cell enhancement of IL-17 production (Geis et al., 2015), activation of the NF-κB pathway (J. M. Kim et al., 2002), generation of pro-tumoural myeloid-derived suppressor cells (Thiele Orberg et al., 2017) and increasing IL-8 secretion (Hwang et al., 2013).

The *B fragilis* toxin is also linked to DNA damage via the induction of spermine oxidase (SMO), a polyamine catabolic enzyme that is highly inducible by inflammation. SMO produces reactive oxygen species such as superoxide and hydrogen peroxide, which in turn result in induction of stress response pathways and DNA damage. Intestinal SMO expression is elevated in mice infected with ETBF and, conversely, inhibition of SMO in ETBF-infected mice reduces chronic inflammation, decreases immune response upregulation and inhibits colon tumourigenesis (Goodwin et al., 2011).

All these characteristics implicate ETBF in the early stages of colon carcinogenesis. Therefore, unsurprisingly, ETBF have been found more often in CRC patients than in healthy controls by multiple groups in multiple countries, providing a strong association of ETBF with CRC (J. I. Keenan et al., 2016; Merino et al., 2011; Ramamurthy et al., 2013; Toprak et al., 2006). Our findings support a direct role for ETBF in carcinogenesis (Purcell et al., 2017), and further evidence for ETBF involvement in colorectal carcinogenesis is found in the ability of ETBF to substantially increase the rate at which $APC^{Min+/-}$ mice develop tumours in their colons, compared to uninfected mice (Rhee et al., 2009). In the same study, ETBF-infected mice with wild type APC, while not developing tumours, developed inflammation of the colon. The finding that colonic carriage of ETBF causes colonic inflammation in normal mice and colonic tumours in mice with mutant *APC*, coupled with multiple findings of ETBF being positively associated with CRC in humans, suggest ETBF may both have a role in initiating carcinogenesis and be able to persist in the developing tumour environment, consistent with the alpha-bug model (Sears & Pardoll, 2011) described in Section 1.5.1.

It should be noted however that asymptomatic colonic carriage of ETBF is common (J. I. Keenan et al., 2016; Odamaki et al., 2012). While it is currently unknown whether

asymptomatic carriers display any colonic inflammatory pathology, this suggests that ETBF-mediated colon carcinogenesis occurs in a minority of carriers, and other environmental or lifestyle factors must be required.

1.5.2 Lifestyle risk factors for CRC

Multiple lifestyle factors may contribute to colorectal carcinogenesis; red meat consumption (Chan et al., 2011), smoking (Botteri et al., 2008), alcohol (Fedirko et al., 2011), obesity (Larsson & Wolk, 2007) and lack of physical activity (Pan & DesMeules, 2009) have all been associated with increased risk of CRC. However, elucidating the molecular mechanisms of individual risk factors is difficult as often multiple lifestyle factors such as smoking, alcohol and chronic inflammation are found in the same patient (Haas et al., 2012).

Chronic alcohol consumption can result in nutritional deficiencies with reduced vitamins B1, B2 and B12 resulting in an increase in reactive oxygen species (Testino, 2011), while alcohol-induced reduction in folic acid can alter the production of S-adenosyl methionine (SAM), a key methyl donor, leading to alterations in the epigenetic control of gene expression (Sauer et al., 2010).

Tobacco smoke contains numerous carcinogenic compounds that can affect multiple molecular pathways (Derry et al., 2013). In addition to conferring an additional risk for CRC, there is also evidence to suggest that smoking may be associated with a higher risk of proximal rather than distal cancer (Botteri et al., 2008). Smoking is also linked to microsatellite instability, CpG island methylator phenotype and BRAF mutant tumours (Limsui et al., 2010). This connection of smoking with the microsatellite instability pathway (Section 1.1.2) indicates smoking-induced epigenetic changes may underlie smoking-associated CRC. Of particular interest, a recent study showed an association of smoking with hypermethylation of the *APC* promoter in colorectal cancer (Barrow et al., 2017).

Obesity is another established risk factor for CRC (Moghaddam et al., 2007), particularly in men but with a lesser effect in women (Bardou et al., 2013). The mechanisms behind this association remain to be fully elucidated, but the association of chronic inflammation and obesity may increase the risk of CRC (John et al., 2007), and gut microbes may be involved in promoting inflammation in people with obesity (Cani et al., 2012). Additionally there is a strong link between diet and obesity, with diets high in rapid carbohydrates and fat being linked

to CRC risk in patients with obesity, while dietary fibre and polyunsaturated fats from fish appear protective (Johnson & Lund, 2007).

The assertion that high levels of fibre in the diet may reduce the risk of CRC was postulated as far back as the 1970s, based on low rates of CRC in rural Africans who ate a high-fibre diet (Burkitt et al., 1974). More recent analyses have confirmed the reduced risk of CRC with dietary fibre intake (Aune et al., 2011).

Another dietary risk factor for CRC is red meat intake, with high red meat consuming countries, including New Zealand having higher rates of CRC than countries with lower red meat intake (Chan et al., 2011). Although the mechanism(s) that increase CRC risk from red meat consumption have not been elucidated, the presence of heme iron, resulting in the production of nitroso compounds, has been postulated due to the lack of increased CRC risk in populations that eat poultry and other meats that do not contain heme iron (Bingham et al., 2002).

1.5.3 Interplay of diet, gut bacteria and metabolites in CRC

There is a growing awareness that the unique mix of different bacterial species have the potential to contribute to an individual's risk of diseases through their breakdown of undigested food in the colon into a broad range of dietary metabolites (Louis et al., 2014). An individual eating a diet rich in fruit and vegetables is more likely to produce increased levels of butyrate, but when the same individual (colonised with the same gut microbiota) switches to a high fat, carbohydrate-rich diet, the types of dietary metabolites produced in the colon also change (O'Keefe et al., 2015).

Fibre is fermented in the gut to produce short chain fatty acids such as butyrate and propionate by colonic bacteria (Louis & Flint, 2017). The short chain fatty acids may protect against CRC by a number of different mechanisms. Butyrate is an inhibitor of histone deacetylase, thereby maintaining histone acetylation, altering gene expression and arresting cell proliferation (Davie, 2003). Additionally, butyrate suppresses colonic inflammation by inhibition of the interferon-gamma and STAT1 pathways (Klampfer et al., 2003; Stempelj et al., 2007).

In addition to the anti-tumourigenic properties described above, butyrate increases mucin production in colonic epithelial cells, strengthening the mucus layer of the colon, enhancing the adherence of beneficial Lactobacilli and Bifidobacteria, and helping to exclude pathogenic bacteria (Jung et al., 2015). This gains significance with evidence of a butyrate gradient in the

colon, with higher levels in the more proximal regions of the colon that reflect saccharolytic fermentation of available carbohydrate by bacteria at this site (Korpela, 2018).

Under normal conditions protein fermentation mostly occurs in the distal colon. Many bacterial proteases are sensitive to acidic pH, and the more neutral pH of the distal colon favours digestion of protein at this site (Macfarlane et al., 1988). When present, dietary fibre decreases the requirement for amino acids as an energy source and reducing the pH through production of short chain fatty acids, thereby reducing the activity of bacterial proteases (Smith & Macfarlane, 1998). Conversely, diets lacking fermentable carbohydrate or diets high in protein can increase both the quantity and location of proteolytic fermentation in the colon (Diether & Willing, 2019; Korpela, 2018).

The switch from fibre to protein fermentation is associated with a change in the relative abundance of fibre-degrading versus protein-degrading bacterial species of colonic bacteria. This is illustrated by one study, where human volunteers given an exclusively animal-based diet showed increased abundance of proteolytic bacteria, while those given an exclusively plant-based diet increased the abundance of species capable of metabolising plant polysaccharides (David et al., 2014).

The significance of predominantly proteolytic fermentation in the colon is the increased production of metabolites with pro-oncogenic effects. Among these biologically active metabolites, hydrogen sulphide and nitroso compounds are toxic to intestinal cells and have been implicated in the development of CRC (Hughes et al., 2000). A further metabolite of proteolytic fermentation is p-cresol, a genotoxic compound that can cause DNA damage in colonocytes (Andriamihaja et al., 2015).

Interestingly, proteolytic fermentation occurring primarily in the distal colon correlates with a higher incidence of CRC tumours in the distal colon, a particularly notable characteristic of early-onset CRC as described below, perhaps providing evidence for a link between diet, microbes and early-onset CRC.

1.6 Early onset colorectal cancer

In line with a worldwide increase in incidence of early-onset colorectal cancer (EOCRC) (Bailey et al., 2015; Hessami Arani & Kerachian, 2017), rates of the disease in patients under

50 years of age in New Zealand are increasing in contrast to an overall decline in CRC in recent years (Gandhi et al., 2017).

In addition to the familial conditions mentioned above, other factors that may play a role in CRC in the young include inflammatory bowel disease, which confers a two- to three-fold increase in risk of CRC (Triantafillidis et al., 2009), and prior radiation therapy, such as for paediatric malignancies (Hill et al., 2007). However, the majority of early-onset CRC cases are sporadic (Connell et al., 2017) and this may, in part, reflect changing lifestyles (see Section 1.7).

Patients with EOCRC are not included in early screening initiatives that internationally start at ages between 50 and 65 years. Younger patients are less likely to seek medical advice for gut issues (Bleyer, 2009) and, when they do, often do not meet criteria for colonoscopy that might lead to a diagnosis. These factors partly explain why EOCRC patients more commonly present with symptomatic and late-stage cancers with poorly differentiated cancers when compared to CRC as a whole (26.3% vs 20%) (Chou et al., 2011; Fleming et al., 2012).

The normal progression of colorectal cancer through the accumulation of multiple genetic abnormalities is a slow process, potentially taking decades. This has led to sporadic EOCRC being considered a specific subtype of CRC with distinct clinical and molecular features (Kirzin et al., 2014; Perea et al., 2014; Raman et al., 2014). Histologically, EOCRC cases more frequently present with mucinous tumours than older-onset CRCs, and a higher percentage of EOCRC tumours have presence of signet-ring cells than older-onset CRCs (Silla et al., 2014), both features that are common in hereditary cancers. Despite the majority of EOCRC patients not being found to have an inherited disorder, they are more likely than older patients to have a family history (Kirzin et al., 2014).

The majority of EOCRC tumours do not show microsatellite instability and those that do are usually due to HNPCC. Therefore microsatellite instability in sporadic EOCRCs is relatively rare (Ballester et al., 2016). The V600E mutation in the *BRAF* gene is less common in EOCRC, with a large multi-cohort study finding increasing prevalence of *BRAF* V600E in older age groups, from less than 4% in under 30 year olds to 13% in over 70 year olds (Willauer et al., 2019). The CpG island methylator phenotype is also found less in EOCRC patients than in older-onset CRC patients, with one study reporting the CIMP to be absent in all 47 EOCRCs, but present in 15/49 (31%) of CRCs in over 60 year olds (Kirzin et al., 2014).

Three decades ago biological differences were identified in tumours located in either the proximal (right-sided) or distal (left-sided) colon, leading to the proposal of distinct categories of colorectal cancer based on tumour location (Bufill, 1990). In particular, proximal tumours had characteristics similar to HNPCC, while left-sided tumours displayed characteristics more similar to FAP. EOCRC patients present more often with distal tumours than older CRC patients. One study had 32% of patients aged 35-39 diagnosed with having tumours in their rectum, with the percentage in subsequent age groups decreasing to 15, with 1% in those aged over 85 years. Conversely, only 9.3% of the 35-39 years age group had tumours in the caecum, rising to 23.2% in the over 85 year olds (Davis et al., 2011). Further to Bufill's categorisation of CRC by tumour location, it has been suggested that EOCRC should be similarly categorised by location (Perea et al., 2015).

1.6.1 Colorectal signet-ring cell carcinoma

Colorectal signet-ring cell carcinoma (SRCC) is a rare form of CRC, accounting for approximately 1% of all cases (Kang et al., 2005). However, SRCCs are more common in younger patients, accounting for 3-13% of EOCRCs (Mauri et al., 2019). They are defined as tumours containing at least 50% of cells displaying signet-ring morphology, where large amounts of intracellular mucin push the nuclei to the edge of the cells (Bosman et al., 2010), with the subsequent loss of cell-cell contact increasing the invasiveness and metastatic potential of the disease. Colorectal SRCCs are associated with younger age than conventional CRCs (Foda et al., 2018; Hyngstrom et al., 2012), generally present at a later stage, with a worse tumour grade and a poorer prognosis (Barresi et al., 2017; Borger et al., 2007). Indeed the American Joint Committee on Cancer accept signet-ring histology as an independent prognostic factor for CRC (Compton et al., 2000). Moreover, there is evidence that a signet-ring component of less than 50%, hence not enough to classify a tumour as SRCC, still has a negative effect on prognosis (Inamura et al., 2015; Tan et al., 2015). Colorectal SRCCs are, like EOCRCs in general, also predominantly found in the distal colon, particularly the sigmoid colon and rectum (Chang et al., 2012; Hyngstrom et al., 2012).

The relative rarity of colorectal signet-ring cell carcinomas is reflected in a paucity of molecular studies, and those that have been published generally report a small number of cases. Accordingly, there are often conflicting conclusions regarding the characteristics of colorectal SRCC. This is illustrated by high rates of microsatellite instability reported in two studies

(Kakar & Smyrk, 2005; Ogino et al., 2006), whereas a third study found all colorectal SRCC samples were microsatellite stable (Nam et al., 2018). Moreover, when present, there is no evidence that microsatellite instability in colorectal SRCC confers an improved prognosis as reported for conventional colorectal adenocarcinoma (Kakar & Smyrk, 2005).

As with HDGC, loss of E-cadherin is frequently found in colorectal SRCC (Borger et al., 2007; Wang et al., 2016). However, unlike HDGC, the presence of *CDH1* mutations, or their role in the pathogenesis of colorectal SRCC, have not been established. E-cadherin loss, and subsequent loss of cell-cell interaction and morphology, may explain the increased invasiveness and metastatic potential of colorectal SRCC (Borger et al., 2007).

1.7 Lifestyle factors in EOCRC

Sporadic colorectal cancers generally emerge from an accumulation of genetic alterations that accumulate over many decades. However, by their very nature EOCRC cases develop in a shorter timeframe. It is likely that the development of tumours in this younger cohort is accelerated by the effects of lifestyle factors. While finding that the risk factors for CRC detailed above (Section 1.5.2), such as obesity and smoking, were also relevant for EOCRC, a recent study comparing patients diagnosed with CRC under the age of 50 with those aged over 50, identified non-modifiable risk factors, such as sex, race, inflammatory bowel disease and a family history of CRC, distinguished EOCRC cases from late-onset CRC (Gausman et al., 2019). Patients with EOCRC were more likely to be male and there were higher rates of EOCRC amongst African Americans and Asians. This reflects an earlier large-scale study of the Surveillance, Epidemiology and End Results (SEER) database that found higher rates in African Americans and Asians aged under 50 living in the USA, despite lower overall CRC rates for Asians (Rahman et al., 2015). This is in line with increasing rates of CRC in Asia, which is also hypothesised to be due to an increased uptake of a Western diet (Y. Deng, 2017).

Similarly, a recent prospective study of young women found the risk of early-onset CRC in women with obesity (BMI>30) to be nearly double that of women with a BMI between 18.5 and 22.9 (Liu et al., 2019), possibly reflecting earlier, and increasing exposure in recent decades, to Western diets and consequent risk factors such as obesity.

Interestingly, in light of the higher risk of EOCRC in African Americans, highlighted by Gausman and colleagues, the large disparity in colon cancer between African Americans and

native Africans from South Africa has previously been noted (O'Keefe et al., 2007), and been associated with a high fat, low fibre diet, with African Americans displaying low levels of short chain fatty acids in their colons (O'Keefe et al., 2015). So, modifiable factors, such as a Western diet and the evidence of nutritional transition associated with an increasing uptake of Western-style diets across the world in recent decades, may play a role in the increasing incidence of EOCRC.

1.8 Cellular models of colorectal carcinogenesis

While advances in genomic technologies, some of which are utilised in this thesis, have allowed a rapid evaluation of clinical samples and helped elucidate the molecular basis of disease, suitable *in vitro* models of disease are still required to further understand the aetiology of cancer, the effects of risk factors and to test possible therapeutic strategies. As such the cancer cell line remains a mainstay of cancer research. Many of the cell lines commonly used to study cancer date back 50 or more years and questions around their clinical relevance have arisen (Borrell, 2010).

The resemblance of a cell line to the original tumour is not always strong. As early as the 1970s concerns were raised that cell lines can acquire additional modifications over time as they are cultured (Nelson-Rees et al., 1976). One commonly used colorectal cancer cell line, Caco-2, displays variability in growth characteristics with increasing passage number. Cells at passage number 72 reached a growth plateau quicker than passage number 22 and 33 cells and, after 21 days growth, cells at passage 72 also had lower alkaline phosphatase expression (Briske-Anderson et al., 1997).

In light of these differences the clinical relevance of cell lines is an issue that needs to be raised when designing *in vitro* studies. An important consideration, is that cancer cell lines, by virtue of being derived from tumours, already have numerous genetic and epigenetic alterations and therefore likely do not represent the normal state of cells within non-neoplastic tissue. This is a particularly important consideration when investigating the early stages of carcinogenesis, where the failure to take into account the effects of stimuli on the cells, in light of the existing mutational background of the cells, may compromise conclusions made in relation to early events in carcinogenesis. This is exemplified by the response of cultured colonic epithelial cell lines to the *B fragilis* toxin.

Cellular approaches to research into ETBF-mediated CRC have primarily been focused on the HT29 cell line and its derivative, HT29/c1 (Huet et al., 1987). HT29 cells display an exquisitely sensitive response to BFT, with the cells rapidly losing their morphology and rounding up in response to toxin-mediated cleavage of E-cadherin (Sears, 2009) (Figure 1.8). This rapid response has led to HT29 being the cell line of choice for studying cellular reaction to the *B. fragilis* toxin (Hwang et al., 2013; J. M. Kim et al., 2002; Sears, 2009). However, our results show that this response is not universal amongst colon cancer cell lines, as seen in the failure of HCT116 cells to round up upon exposure to BFT (Figure 1.8).

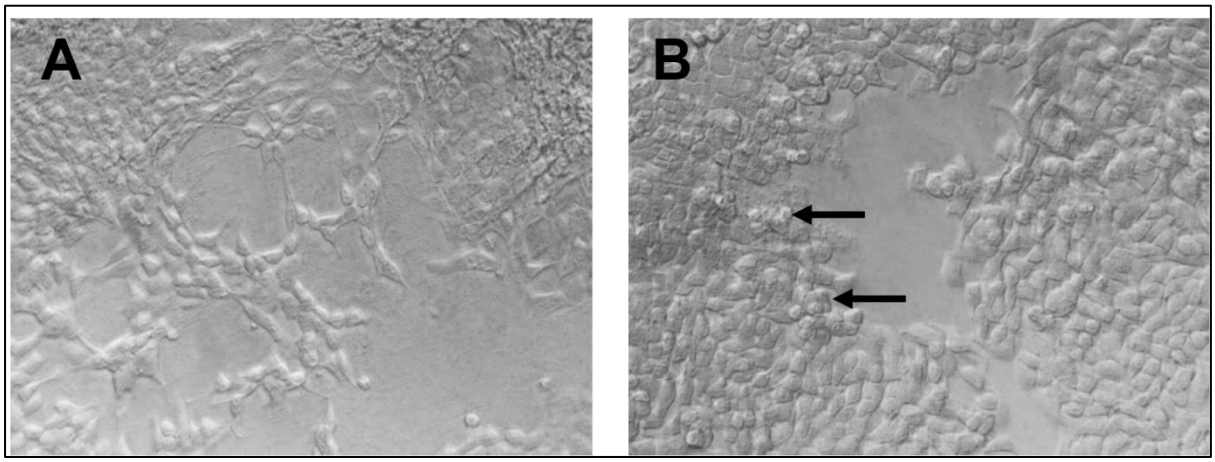

Figure 1.8 Colorectal cancer cell lines HCT116 and HT29 after incubation with *B. fragilis* toxin. A, HCT116 cells maintain their cellular morphology after addition of the toxin while, B, HT29 cells show characteristic rounding (arrowed) (Keenan, unpublished).

The underlying cause of the rapid response by HT29 cells to BFT is likely to be mutations within critical genes that make HT29 cells particularly susceptible to the effects of ETBF. This illustrates the idea that the use of cells that are already highly susceptible to the effects of BFT may not represent the best model for studying any bacterially-induced induction of carcinogenesis.

1.9 APC in HT29 cells

As discussed earlier, the *APC* gene is mutated in the majority of sporadic colorectal cancers. Additionally, Miyoshi and colleagues found that 60% of colorectal tumours they studied with *APC* mutations had two mutations (Miyoshi et al., 1992). Similarly, HT29 cells contain two

truncated forms of the *APC* gene and as a result express no full length APC protein (Morin et al., 1996). Consequently, due to the central role of APC in the formation and maintenance of the cytoskeleton, it is possible that the lack of full-length APC is fundamental to the rounding up of cells, when E-cadherin is cleaved following exposure to BFT.

In light of this, we sought to introduce wildtype *APC* into HT29 cells in order to produce a cell line that would be more informative when studying the effects of environmental factors such as bacterial toxin exposure on the early stages of colorectal carcinogenesis (Chapter 5). Wildtype APC has been expressed in HT29 cells previously using a plasmid containing an inducible *APC* gene (Morin et al., 1996). However, expression was under the control of a zinc-inducible promoter. Unfortunately, given that BFT is a zinc metalloprotease (Moncrief et al., 1995), using zinc to induce APC expression would have created uncertainty over whether observations were due to APC expression or the effects of zinc on BFT.

Therefore we sought to edit an existing allele of *APC* in HT29 cells to correct the mutation to allow the expression of full-length protein. Additionally, the correction of the cell's existing *APC* gene would allow for normal cellular expression of the protein rather than artificial induction of the gene, thereby better representing the natural *in vivo* situation.

The CRISPR-Cas system, consisting of CRISPR (clustered regularly interspersed short palindromic repeats) and CRISPR-associated (Cas) genes is used by bacteria as part of their adaptive immunity. Briefly, viral or plasmid DNA entering a cell are recognised as foreign and digested into small fragments. Some of these small fragments are subsequently incorporated into the CRISPR array of the bacterial genome. The CRISPR array contains small fragments of DNA from previous infections. When the array is transcribed into RNA and subsequently cleaved into individual CRISPR RNAs (crRNAs) each individual crRNA will recognise a specific invading virus or plasmid and guide Cas nucleases that are complexed to the crRNA to digest the invading DNA upon re-exposure (Barrangou et al., 2007).

Jinek and colleagues initially showed that the Cas9 nuclease could be programmed to target double-stranded DNA cleavage at specific DNA sites *in vitro* by the use of crRNAs with specific target sequences, known as guide RNAs (gRNAs) (Jinek et al., 2012). This CRISPR/Cas9 technique was subsequently used to edit genomes in mammalian cells (Cong et al., 2013; Mali et al., 2013). Also, in another innovation a nickase enzyme, Cas9n, was used to nick both strands of DNA at locations close to each other, using separate gRNA recognition sequences for each strand in order to minimise off-target effects, as both sites require nicking

for homology-directed repair to occur (Cong et al., 2013). The CRISPR/Cas9n system allows for the precise editing of mutated bases such as those found in the *APC* gene of HT29 cells.

1.10 Aims of the study

This study is divided into three parts: firstly determining the status of E-cadherin and *CDH1* mutation in a cohort of young colorectal SRCC patients; secondly investigating the *CDH1* and *APC* mutation status of a New Zealand cohort of EOCRC patients, and finally correcting mutant *APC* in a CRC cell line to obtain a model for the study of factors initiating colorectal carcinogenesis.

The first part of the study utilised a cohort of Pakistani early-onset colorectal SRCC samples, to look at the mutation status of the *CDH1* gene using a next generation sequencing approach. This work was devised based on the similarity between colorectal SRCC and the presence of SRCC in hereditary diffuse gastric cancer, and the common finding of inherited *CDH1* mutations in HDGC. This was complemented by immunohistochemistry to identify expression of E-cadherin and mismatch repair genes. The results of this work are presented in Chapter 3.

The second part of the study builds on Chapter 3 by looking at *CDH1* mutation status in an archival New Zealand cohort of EOCRC patients, primarily with adenocarcinomas, using the same approach as in Chapter 3. Additionally we devised a similar approach to look at *APC* mutation within the same samples. Again we looked at E-cadherin expression in these samples, while their mismatch repair gene expression status was obtained from the pathology reports. Results obtained from this work were correlated with clinical data and are presented in Chapter 4.

The final part of the study used a CRISPR-Cas9n gene editing approach to correct one allele of the *APC* gene in HT29 cells. Correction of this gene will allow the expression of wild-type APC protein in these cells. The resultant cell line was then characterised in comparison to the parent HT29 cell line, including observation of its response to BFT. The results of these experiments are presented in Chapter 5.

Chapter 2 Materials & Methods

2.1 Materials

Key reagents and kits used in this research are listed in the tables below.

Table 2.1 Primary antibodies

Antibody	Raised In	Product code	Supplier	Analysis
E-cadherin	Mouse	NCH-38	Dako (Glostrup, Denmark)	IHC
MSH6	Rabbit	SP93	Roche (Basel, Switzerland)	IHC
PMS2	Mouse	A16-4	Roche	IHC
MLH1	Mouse	M1	Roche	IHC
E-cadherin	Mouse	ab76055	AbCam (Cambridge, UK)	IF
β-catenin	Mouse	ab22656	AbCam	IF
APC (N-terminal)	Mouse	ab58	AbCam	WB
APC (C-terminal)	Rabbit	ab15270	AbCam	WB

IHC, Immunohistochemistry; IF, Immunofluorescence microscopy; WB, Western blotting.

Table 2.2 Secondary antibodies

Antibody	Product code	Supplier	Analysis
Alexafluor 488-conjugated goat anti-mouse	A11001	Invitrogen (Carlsbad, CA, USA)	IF
Alexafluor 488-conjugated goat anti-rabbit	A11008	Invitrogen	IF
Alexafluor 594-conjugated donkey anti-mouse	A21203	Invitrogen	IF
Horseradish peroxidase-conjugated goat anti-mouse	P 0447	DAKO	WB
Horseradish peroxidase-conjugated goat anti-rabbit	P 0448	DAKO	WB

IF, Immunofluorescence microscopy: WB, Western blotting.

Table 2.3 Kits used in this study

Name	Product Code	Supplier
QIAamp DNA FFPE Tissue Kit	56404	Qiagen (Hilden, Germany)
DNeasy Blood & Tissue Kit	69504	Qiagen
MiSeq 500-Cycle Reagent Kit	MS-102-2003	Illumina (San Diego, CA, USA)
NucleoBond Xtra Midi Kit	740410.50	Macherey-Nagel (Düren, Germany)
PureLink Quick Gel Extraction Kit	K2100-12	Invitrogen
Qubit dsDNA High Sensitivity Assay Kit	Q23854	Thermo Fisher (Waltham, MA, USA)

Table 2.4 Cell processing and staining reagents

Name	Product Code	Supplier	Used for
Cell Conditioning 1 retrieval solution	950-124	Roche Ventana (Oro Valley, AZ, USA)	IHC
UltraView DAB	760-500	Roche Ventana	IHC
Hematoxylin II	790-2208	Roche Ventana	IHC
Bluing Reagent	760-2037	Roche Ventana	IHC
Trypan blue	T6146	Sigma (St Louis, MO, USA)	
Texas Red-X phalloidin	T7471	Invitrogen	IF
Hoechst 33342	H3570	Invitrogen	IF
Antifade mountant	P36930	Molecular Probes (Eugene, OR, USA)	IF

IHC, Immunohistochemistry; IF, Immunofluorescence microscopy

Table 2.5 Bacterial and cell culture reagents

Name	Product Code	Supplier
McCoy's 5A medium	16600108	Life Technologies (Carlsbad, CA, USA)
Fetal bovine serum	16000044	Life Technologies
Penicillin/streptomycin solution (100x)	10378016	Life Technologies
TrypLE Express Trypsin solution	12605010	Invitrogen
Luria Bertani media	CM0996B	Oxoid (Basingstoke, UK)
Brain Heart Infusion media	CM1135	Oxoid

Table 2.6 Molecular reagents

Name	Product Code	Supplier
Kapa HiFi HotStart Readymix	KK2602	Kapa Biosystems (Basel, Switzerland)
SYBRSafe Gel Stain	S33102	Invitrogen
HighPrep magnetic beads	AC-60005	MagBio (Gaithersburg, MD, USA)
Phusion High Fidelity Polymerase	F530S	ThermoFisher
T4 DNA ligase	15224090	ThermoFisher
Hyperladder DNA size marker	BIO33029	Bioline (Memphis, TN, USA)
XhoI restriction endonuclease	R0146S	New England Biolabs (Ipswich, MA, USA)
XbaI restriction endonuclease	R0145S	New England Biolabs
BpiI restriction endonuclease	R0539S	New England Biolabs
KpnI restriction endonuclease	R0142S	New England Biolabs
NotI restriction endonuclease	R0189S	New England Biolabs

Table 2.7 Western blotting reagents

Name	Product Code	Supplier
Pierce ECL Plus Western Blotting Substrate	32132	Thermo Fisher
Amersham Hybond-P polyvinylidene fluoride (PVDF) membrane	10600023	GE Healthcare (Chicago, IL, USA)
Mini-PROTEAN TGX pre-cast SDS-PAGE gels	456-1094	BioRad (Hercules, CA, USA)

2.2 Methods

2.2.1 Human tissue samples

2.2.1.1 DNA extraction from tissue

DNA was extracted from up to three 10 µm sections of formalin-fixed paraffin-embedded samples using a QIAamp DNA FFPE Tissue Kit (Qiagen) according to the manufacturer's protocol. Briefly, sections were placed in a 1.5 mL microcentrifuge tube and deparaffinised by the addition of 1 mL xylene with vigorous vortexing. Following centrifugation at 13,000 x *g* for 2 minutes, the xylene was removed and 1 mL ethanol added to remove residual xylene. After centrifugation at 13,000 x *g* for 2 minutes, removal of ethanol and air-drying of the pellet, lysis buffer and proteinase K were added and samples were incubated at 56°C for 1 hour. Further incubation at 90°C for 1 hour was done to reverse formaldehyde modification of DNA. Subsequently, 200 µL protein precipitation buffer and 200µl ethanol were added and mixed. The mixture was then centrifuged through a spin column, washed twice with wash buffers and eluted using a Tris-EDTA solution supplied with the kit.

To extract DNA from frozen samples, up to 25 mg of tissue was added to 180 µL lysis buffer (Qiagen DNeasy Blood and Tissue kit) and homogenized in a Precellys Evolution homogenizer (Bertin Technologies, Montigny-le-Bretonneux, France) using 2.8 mm zirconium oxide beads in reinforced 2 mL microtubes for 4 x 15 s at 5800 rpm. DNA was then extracted according to the protocol provided with the kit. Briefly, 20 µL of proteinase K was added to the homogenised sample, mixed and incubated at 56°C overnight. Two hundred µL of protein precipitation buffer and 200 µL ethanol were then added and mixed. The mixture was centrifuged through a spin column, washed twice with wash buffers and eluted using a Tris-EDTA solution.

2.2.1.2 Histology

Hematoxylin and eosin (H&E) stained sections were evaluated according to the WHO criteria. Adjacent sections were immunostained to identify the presence of E-cadherin, and to identify microsatellite markers MSH6, PMS2 and MLH1. Staining was carried out using the Roche Ventana BenchMark Ultra Platform stainer. Briefly, antigen retrieval for E-cadherin was for 64 minutes at 95˚C, and for MSH6 was for 32 minutes at 100˚C. For MLH1 and PMS2, slides were immersed in Roche Ventana Cell Conditioning 1 (CC1) retrieval solution for 64 minutes

or 92 minutes respectively. Incubation in primary antibody was at 37˚C for 40 minutes (E-cadherin), or 36˚C for 12 minutes (MSH6), 24 minutes (MLH1) or 32 minutes (PMS2). All sections were subsequently treated with Ventana UltraView DAB to visualise primary antibody staining. Counterstaining was with Roche Ventana Haematoxylin II (modified Mayer's) for four minutes before the sections were blued with Roche Ventana Bluing Reagent, four minutes. E-cadherin staining of signet ring cells was graded as negative, weak positive or positive whereas microsatellite markers were assessed as positive or negative. All slides were scored by a pathologist blinded to the identity of the samples.

2.2.2 Human cell lines

2.2.2.1 Cell culture

Two colonic epithelial cell lines, HT29 (ATCC HTB-38) and HCT116 (ATCC CCL-247) were used in this work. Both are adherent cell lines, isolated from a 44 year old Caucasian patient who presented with colorectal adenocarcinoma (Fogh, 1975) and an adult male diagnosed with colorectal carcinoma (Brattain et al., 1981), respectively.

Both cell lines were grown in McCoy's 5A medium supplemented with 10% (v/v) fetal bovine serum and 100 U/mL penicillin and 100 µg/mL streptomycin. The cells were cultured at 37°C and 5% CO_2 in a humidified incubator, with the media changed every 72-96 hours. Cells nearing confluency were washed twice with PBS and lifted from the flask with 1 mL 0.25% TrypLE Express at 37°C and 5% CO_2 for 10 minutes. Trypsinisation was stopped by the addition of 9 mL of medium and cells were centrifuged at 1500 x *g* for 5 minutes. The supernatant was removed and cells were either resuspended in fresh medium, or used for experimentation.

2.2.2.2 Counting

To count cells, 15 µL were taken from 10 mL of trypsinised cell suspension (see above) and added to 15 µL of trypan blue solution (0.25% w/v in PBS). Cell numbers were quantified using a haemocytometer. Dead cells stain blue with trypan blue, enabling the number of non-viable cells to be calculated as a proportion of the total cell count (Chae, 2017).

2.2.2.3 Preservation

For long term storage, cells were centrifuged at 1500 x *g* and the supernatant removed. Following a brief PBS wash, pelleted cells were resuspended in freezing medium (90% FBS, 10% DMSO) to give a concentration of 1.5-3 x 10^6 / mL. One milliliter aliquots were frozen overnight in cryotubes at -80°C and then transferred to liquid nitrogen.

2.2.2.4 Light microscopy

Images of cells cultured in tissue culture plates were captured using an Olympus CK2 microscope equipped with an Olympus DP21 digital camera (Olympus, Shinjuku, Japan).

2.2.2.5 Immunofluorescence microscopy

Cells were cultured on sterile glass coverslips (Paul Marienfeld GmbH, Lauda-Konigshofen, Germany) before being fixed in a pre-chilled (-20°C) 50:50 (v/v) solution of methanol:acetone for 45 minutes at 4°C. The cells were then washed 4 x 5 minutes with PIPES buffer (Appendix 2) at room temperature before permeabilisation with PIPES buffer containing 0.5% (v/v) Triton X-100 for 15 minutes at room temperature. Following two further 5 minute washes in PIPES buffer, non-specific binding sites were blocked with 1% BSA in PIPES buffer for 45 minutes at room temperature with gentle agitation. Coverslips were then washed a further 2 x 5 minutes in PIPES buffer before being incubated in primary E-cadherin antibody (diluted 1:250 in PIPES buffer with 1% BSA) overnight at 4°C. Following four washes with PIPES buffer at room temperature, the coverslips were incubated for 90 minutes at room temperature in secondary antibody (Alexafluor conjugated secondary antibody diluted 1:1000 in PIPES with 1% BSA) before being washed four times with PIPES.

Actin was detected using Texas Red-X phalloidin (5 µL/mL) by incubating overnight at 4°C in PIPES with 1% BSA. Nuclei were stained using Hoechst 33342 (Invitrogen) at 2.5 µg/mL in PIPES buffer for 30 minutes at room temperature. Coverslips were washed a further two times in PIPES before being mounted onto slides using antifade mountant (Molecular Probes).

Fluorescent imaging was carried out using an AxioImager Z1 microscope (Zeiss, Oberkochen, Germany) with EC Plan-Neofluar lenses (Zeiss) and an Axiocam 506 color camera (Zeiss). Images were captured by Zen3 software (Zeiss). Exposure times for each channel (red, blue,

green) were kept constant in all images from each experiment to maintain an accurate comparison between variables.

2.2.3 Next generation sequencing of *CDH1* and *APC*

2.2.3.1 *CDH1* sequencing library construction

All coding regions and the proximal promoter of the *CDH1* gene were amplified using a two-step PCR strategy previously described by Dr Chris Hakkaart (Hakkaart et al., 2019). Briefly, the targeted regions were amplified initially using primers specific to the region of interest, with 18 base pairs (bp) non-specific tails at 5' ends, which acted as binding sequences for primers in the second round of PCRs (Figure 2.1). Seventeen different PCR reactions were required to provide coverage of all 16 exons and the proximal promoter sequence of *CDH1*. Each amplicon was designed to be less than 500bp to fit within the capabilities of a MiSeq 500-cycle reagent kit (Illumina). Consideration was also given to the relative composition of nucleotides in the first four bases across all primers as during a MiSeq run the first four bases are used for fluorescence calibration. Thus, a bias toward any particular nucleotide at each base would result in suboptimal calibration. Primer sequences are detailed in Appendix 1.

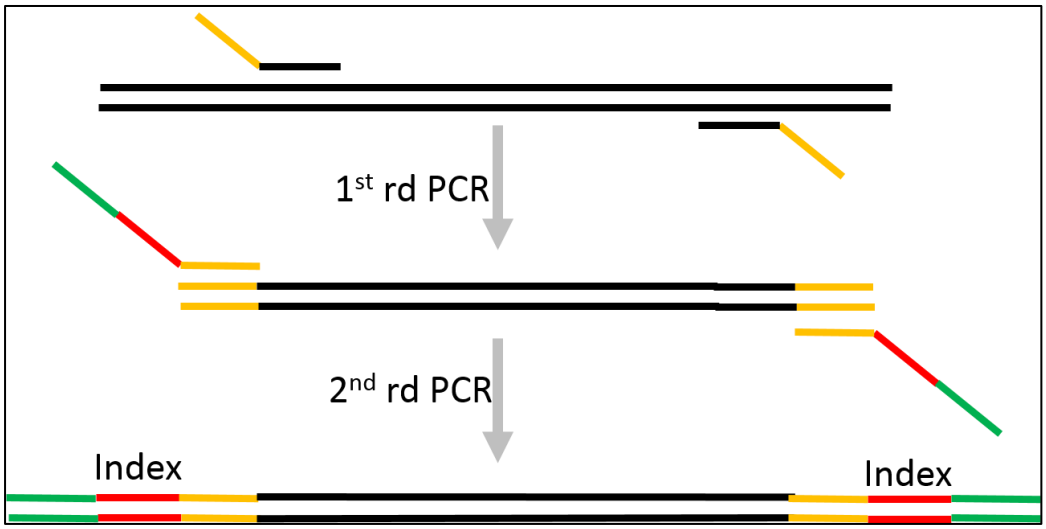

Figure 2.1 Two-step PCR strategy to generate amplicon sequencing libraries. A first round of PCR is carried out using locus-specific primers with 18bp 5' ends as templates for second round PCR. The second round primers, used on pooled first round products for each sample, contain unique paired indices allowing identification of samples and binding sites for sequencing primers used by the Illumina MiSeq.

2.2.3.2 *APC* sequencing library construction

PCR primers were designed to create amplicons spanning regions of the *APC* gene accounting for the vast majority of known mutations found in cancers. Regions of interest were identified using the COSMIC database of somatic mutations in cancer (https://cancer.sanger.ac.uk/cosmic). Fifteen amplicons were designed spanning multiple regions of the *APC* gene including the mutation cluster region. The same 18bp non-specific 5' ends used for the *CDH1* first round amplification primers were added to the sequences of *APC*-specific first round primers. Primer sequences and target nucleotides are detailed in Appendix 1.

2.2.3.3 Primer design

Where PCR was not performed using previously published primers, primer design was carried out using the Whitehead Institute web design page Primer3 (primer3.ut.ee/). Off-target binding was checked using the BLAST tool from the US National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/tools/primer-blast/).

All primers were ordered from Integrated DNA Technologies Inc. (Singapore) and reconstituted to a concentration of 50 µM with Tris-EDTA (TE; Appendix 2). Working stocks were prepared by further dilution to 10 µM in H_2O.

2.2.3.4 First round amplicon generation

Regions of interest within *CDH1* and *APC* were amplified using amplicon-specific primers in 10 µL reactions using Kapa HiFi HotStart Readymix (Kapa Biosystems) to maintain sequence fidelity. The reaction mixture is detailed in Table 2.8.

Table 2.8 Mixture for amplicon-specific PCR

Reagent	Volume
10 µM forward primer	0.6 µL
10 µM reverse primer	0.6 µL
Kapa HiFi HotStart Readymix	5 µL
Sterile MPW	1.8 µL
DNA (15 ng)	2µL
Total	10µL

Amplification of PCR products was carried out in a Kyratec Supercycler thermal cycler (Kyratec, Mansfield, QLD, Australia) using the conditions detailed in Table 2.9. 3 µL of each amplicon was run on a 1.5% agarose gel to check for successful amplification (Section 2.2.3.5). The amplicons for each sample were then pooled and cleaned using HighPrep magnetic beads (MagBio) (see Section 2.2.3.6).

Table 2.9 Thermocycler conditions for amplicon-specific PCR

Step	Temp (°C)	Time	Cycles
Initial Denaturation	95	3 min	1
Denaturation	98	14 sec	35
Annealing	64	14 sec	
Extension	72	14 sec	
Final Extension	72	1 min	1

2.2.3.5 Agarose gel electrophoresis

PCR products were resolved in 1.5% agarose gels prepared by dissolving agarose in 1x Tris Borate EDTA (TBE) (Appendix 2) and supplemented with 0.4 x SYBRSafe. Appropriate volumes of 6 x loading buffer (Appendix 2) were added to samples prior to running gels in 1 x TBE at 100V for up to an hour. Two µL of 100bp Hyperladder DNA marker were run alongside samples as a size marker. Gels were visualized under ultraviolet light using a Q9 Alliance Advanced transilluminator with Q9 Alliance software (Uvitec, Cambridge, UK).

2.2.3.6 Magnetic bead clean-up of PCR products

Pooled PCR products from first or second round amplification steps for sequencing library construction were cleaned using HighPrep magnetic beads to remove unused primers, excess dinucleotides and other contaminants. Samples were mixed with 1.8 x volumes of bead solution by pipetting up and down at least 10 times and left at room temperature for 5 minutes. Samples were placed on a magnet and the beads allowed to aggregate on the walls of the tubes. Supernatants were discarded and the bead aggregates washed twice with 200µl 70% ethanol. The beads were air dried for 15-20 minutes. Sample were removed from the magnet, and the

DNA eluted from the beads with 40µl of the elution buffer supplied with the kit. Samples were replaced on the magnet to aggregate the beads and the supernatants containing the cleaned DNA were recovered to clean tubes.

2.2.3.7 DNA quantitation by Nanodrop

DNA and PCR products were quantified by a Nanodrop 1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). TE, mqH_2O or the respective elution buffer was used as a blank depending on the sample. One µL of sample was placed on the Nanodrop pedestal and the ratio between absorbances at 260 and 280nm used to determine the concentration and purity of the DNA.

2.2.3.8 Adapter PCR

Pooled and cleaned amplicons for each sample were amplified in a second reaction using primers with 18bp 3' sequences complementary to the 18bp 5' ends of the first round amplicons. The second round adapter primers also contain unique indices that identify sequences from individual samples in a pooled library, as well as sequences required to bind the amplicon to Illumina flow cells, and sequences for the binding of Illumina MiSeq extension primers (Figure 2.1). The same pairs of indexed primers were used for *APC* as for *CDH1* for samples in which both genes were sequenced (NZ EOCRC cohort, Chapter 4). The reaction mixture is detailed in Table 2.10. Second round primer sequences are detailed in Appendix 1.

Table 2.10 Adapter PCR mixture

Reagent	Volume
10 µM forward primer	0.6 µL
10 µM reverse primer	0.6 µL
Kapa HiFi HotStart Readymix	5 µL
Sterile MPW	1.8 µL
Pooled 1st rd DNA (2 ng)	2 µL
Total	10 µL

PCR conditions are detailed in Table 2.11. Adapter combinations were unique to each sample, allowing identification of sample for each sequence produced. It was ensured that there were

no similarities in adapter primers between tumour and normal tissue DNA for the same patient (NZ EOCRC cohort) to avoid cross-contamination and ensure that mutations common to both tumour and normal tissue were real.

Table 2.11 Thermocycler conditions for adapter PCR

Step	Temp (°C)	Time	Cycles
Initial Denaturation	95	2 min	1
Denaturation	98	14 sec	
Annealing	68	14 sec	10
Extension	72	14 sec	
Final Extension	72	1 min	1

Three µL of each reaction were run on a 1.5% agarose gel alongside the pre-amplification pooled input DNA to visualize the successful addition of adapter primers to the amplicons, evidenced by an increase in the size of the amplicons. Five µL of each adapter PCR product were pooled as a library in a single tube and cleaned by magnetic bead clean-up (Section 2.2.3.6).

2.2.3.9 Library preparation and sequencing in Illumina MiSeq

Quantitation of sequencing libraries and sequencing of the libraries using MiSeq (Illumina, San Diego, CA, USA) were carried out by Dr Robert Day (Biochemistry Department, University of Otago. Dunedin), as previously described (Hakkaart et al., 2019).

2.2.3.10 Quantitation of sequencing libraries

Sequencing libraries were quantified using a Qubit dsDNA High Sensitivity Assay Kit prior to loading on an Illumina flow cell for sequencing. Qubit working solution was prepared freshly as a 1:200 dilution of Qubit reagent in Qubit buffer. Two µL of sample were added to 198 µL of working solution. Standards were made by adding 190 µL of working solution to 10 µL of each standard. Standards were run alongside samples to calibrate the Qubit fluorometer.

2.2.3.11. Sequencing, analysis and annotation

DNA libraries were diluted to 4 nM and run on an Illumina Miseq using V2-500 cycle reagent kits (Illumina). Raw paired end reads were cleaned using Trimmomatic v.0.35 (Bolger et al., 2014) then the cleaned reads were aligned to the human reference genome (GRCh37/hg19) using the Burrows-Wheeler Aligner v.0.7.10 (H. Li & Durbin, 2009). Minimum read depths were 40 reads. Single nucleotide variants, insertions and deletions were called using the Genome Analysis Toolkit (GATK) v.3.6 (McKenna et al., 2010). The effects of variants were predicted using SnpEff v.4.2 (Cingolani et al., 2012) and putative mutations were visualised using the Integrated Genomics Viewer (Robinson et al., 2011). The functional consequences of missense variants were predicted by a variety of online tools; Polyphen2 (Adzhubei et al., 2013) (using both HumDiv and HumVar datasets), SIFT (Kumar et al., 2009), Provean (Choi et al., 2012) and Mutation Assessor (Reva et al., 2011).

2.2.3.12 Sanger sequencing for confirmation of next generation sequencing results

Purified PCR products were prepared for sequencing by combining 1 ng of PCR product per 100bp with 3.2 pmol of primer and mqH_2O in a total volume of 5 µL. Samples were then sent to the Genetic Analysis Service (Department of Anatomy, University of Otago, Dunedin) where they were sequenced on a capillary ABI 3700xl Genetic Analyser (Applied Biosystems) following a sequencing reaction using BigDye chemistry (Applied Biosystems).

2.2.3.13 Loss of heterozygosity at *APC* locus

The *APC* sequences obtained in Section 2.2.3 were used to detect loss of heterozygosity (LOH) by comparing levels of each allele at four single nucleotide polymorphisms (SNPs) within the *APC* gene. SNPs showing heterozygosity, by presence of both alleles, in normal tissue were considered to have LOH if the read count of one allele was reduced in relation to the alternate allele in tumour tissue. A low number of alleles present for the missing allele likely reflected non-tumour cells, such as lymphocytes, within the tumour tissue. The four SNPs analysed were rs2229992, rs351771, rs1801166 and rs41115. If all the informative SNPs within the *APC* gene in a single patient showed loss of an allele in the tumour tissue then the tumour was considered to display LOH at the *APC* locus.

2.2.4 Using CRISPR-Cas9 to edit the human HT29 colorectal cancer cell line

2.2.4.1 Design of gRNA sequences for editing of mutated *APC* in HT29 cells

The *APC* gene in HT29 cells has compound heterozygous mutations, one of which, a c.*2557G>T* transition resulting in a premature stop codon at codon 853, was edited using the CRISPR-Cas9 nickase system.

Two guide RNAs (gRNAs) were designed to target sequences around the mutation using the design tool on the Zhang lab website (crispr.mit.edu). This site has since been shut down although resources are available via the Zhang lab (https://zlab.bio/guide-design-resources). One gRNA overlapped the mutated base while the other was 40-60 bases upstream of the mutated base.

The use of a nickase enzyme to nick individual strands of DNA at two points close to each other, rather than the conventional single site cutting of both strands, was done to increase the specificity of the process by limiting off-target effects as both gRNA molecules would need to bind to complementary sequences close to each other for the editing process to take place. The chances of this happening at non-target sequences is greatly diminished compared to systems requiring only one gRNA to bind as described in Section 1.9.

2.2.4.2 Plasmids and bacteria used for CRISPR-Cas9n experiments

gRNAs were generated in pSpCas9n(BB)-Puro PX642 plasmids (Addgene) containing the Cas9n gene and the puromycin resistance gene. pSpCas9n(BB)-2A-Puro (PX462) V2.0 was a gift from Feng Zhang (Addgene plasmid # 62987; http://n2t.net/addgene:62987; RRID:Addgene_62987) (Ran et al., 2013). pBluescript KS(+) (Stratagene) is a small plasmid that was used as an intermediary for manipulating the template DNA. Subsequently the template DNA was cut out of pBluescript KS(+) and ligated into pCDNA3.1 (Invitrogen), an expression vector carrying the neomycin resistance gene, to be used for the CRISPR-Cas9n editing. All three plasmids and subsequent engineered plasmids were amplified in and purified from *E.coli* DH5α cells.

2.2.4.3 Transformation of plasmids into bacteria

Plasmids were transformed into *E.coli* DH5α bacteria by thawing 100 µL aliquots of chemically competent cells (a kind gift from Dr Paul Pace, made by a method based on the Hanrahan method for preparation and transformation of *Escherichia coli* (Green & Sambrook, 2018)) on ice, adding approximately 10 ng of ligation reaction or 1 ng of plasmid DNA, mixing gently and incubating on ice for 30 minutes. The mixture was heat shocked at 42°C for 45 seconds then returned to ice for 2 minutes. Pre-warmed Luria Bertani (LB) broth (900 µL) was added and the mixture incubated, shaking at 37°C, for 1 hour. Twenty to two hundred µL of the transformation mix were then plated onto LB agar plates and incubated at 37°C overnight. Colonies were subsequently picked, grown in LB broth and plasmids extracted according to Section 2.2.4.4.

2.2.4.4 Plasmid purification

Following transformation, bacteria containing pSpCas9n(BB)-2A-Puro, pCDNA3.1 or pBluescript KS(+) plasmids were grown overnight in LB broth with 100 µg/mL ampicillin and harvested to generate plasmid stocks for further use in the CRISPR-Cas9 editing process. Plasmids were isolated using NucleoBond Xtra kits (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. Briefly, 100 mL overnight cultures were centrifuged at 5000 x *g* for 10 minutes at 4°C and the pellets resuspended in buffer containing RNase A by vortexing. Equal volumes of an alkaline lysis buffer were added and the mixture inverted several times to mix. After a 5 minute incubation, neutralisation buffer was added and the tubes mixed by inverting. The mixture was added to an equilibrated DNA-binding column and allowed to flow through by gravity. Following two washes with wash buffer, the plasmids were eluted with elution buffer, precipitated with isopropanol and centrifuged at 5000 x *g* for 30 minutes at 4°C. The resulting pellets were washed with 70% ethanol, centrifuged and air-dried. The plasmids were then resuspended in TE buffer and stored at -20°C until use.

2.2.4.5 Design of *APC* template fragment used for editing mutant *APC*

To enable CRISPR nickase directed mutagenesis of genomic *APC*, a plasmid was designed to provide a homologous recombination template that contained the relevant portion of *APC* carrying the mutation to be copied into the target genome. Briefly, a fragment of the human

APC gene was amplified from pCMV-Neo-Bam-APC (Morin et al., 1996) and subsequently cloned into pBluescript KS(+) using KpnI and NotI restriction sites incorporated into the primers used to amplify the fragment (Appendix 1). This fragment covered the sequence encompassing the two gRNA recognition sequences and approximately 100 bp of flanking sequence either side of the recognition sites. The PCR mixture included Phusion High Fidelity Polymerase to minimise misincorporation of nucleotides during the reaction. The PCR mixture, thermocycler conditions and primers are detailed in Appendix 1.

The resultant amplicon and the pBluescript KS(+) plasmid were independently digested with KpnI and NotI restriction enzymes. The enzymes were added consecutively as they have different salt requirements for optimal digestion. Briefly, 20 µL DNA were added to 5 µL NEB buffer 1, 0.5 µL 100 x BSA and 1 µL (10 units) KpnI in a total volume of 50 µL and incubated at 37°C overnight. 2.5 µL 2 M NaCl (final concentration = 100 mM) were added along with 1 µL (10 units) NotI and the digests incubated for a further 2 hours at 37°C. Digested plasmid was treated with calf alkaline phosphatase to prevent self-ligation and PCR fragments were run on a 0.8% agarose gel, the relevant bands were cut out and the DNA fragments were purified using a gel extraction kit (Invitrogen).

The cleaned, digested fragment was then ligated into the cleaned, digested pBluescript KS(+) vector by incubating at an insert:vector ratio of 3:1 using 2 µL of 10 x ligase buffer and 1µl of T4 DNA ligase (Thermo Fisher) overnight at 16°C in a total volume of 20 µL. Ligated DNA constructs were then transformed into competent DH5α cells (see Section 2.2.4.3).

Following growth of the transformed cells and extraction of the newly created template plasmid, XbaI and XhoI restriction sites were created within the gRNA recognition sequences that still existed on the template plasmid, so that they would no longer be recognised by the gRNAs, and to provide confirmation that the template mutant *APC* fragment had been copied into the target genome following transfection into HT29 cells. The mutation primers detailed in Appendix 1 introduce the base changes that create the restriction sites without altering the protein sequence. The XbaI site was introduced first, with the post-PCR mixture incubated with DpnI enzyme at 37°C for 1 hour to digest original methylated DNA, thereby removing unmutated plasmid, and to leave nascent unmethylated but mutated DNA. The mixture was transformed into DH5α cells (Section 2.2.4.3), individual colonies were cultured and the plasmid DNA from these were purified (Section 2.2.4.4). Restriction digest analyses identified positive incorporation of the XbaI site and then the XhoI site was introduced in the same

manner. Phusion High Fidelity DNA Polymerase was used to minimise misincorporation of nucleotides. The PCR mixture used for each mutation PCR along with thermocycler conditions are detailed in Appendix 1. Successful introduction of mutated bases was confirmed by Sanger sequencing.

2.2.4.6 Gel extraction and clean-up of digested plasmids and PCR products

DNA fragments run on 0.8% agarose gels were extracted using a PureLink Quick Gel Extraction Kit (Invitrogen) according to the manufacturer's protocol. Briefly, the gel was solubilized at 50°C for 10 minutes in 3:1 (volume:weight) Solubilisation Buffer before 1 gel volume of isopropanol was added and the mixture centrifuged through a spin column at 12,000 x *g* for 1 minute. The column was washed by centrifuging 500 µL Wash Buffer through at 12,000 x *g* for 1 minute before drying the column and eluting with Elution Buffer by centrifugation.

2.2.4.7 Insertion of template fragment into pCDNA3

Following successful mutation of the template fragment in pBluescript KS(+), the fragment was cut out of the plasmid and ligated into pCDNA3, a larger plasmid that carries the neomycin resistance gene. This was to allow selection of HT29 cells that contain the template fragment by using geneticin, a drug that is toxic to mammalian cells but to which resistance is conferred by the neomycin resistance gene. Both pCDNA3 and the pBluescript KS containing the template fragment were digested with KpnI and NotI as previously (Section 2.2.4.5). Digested pCDNA3 was treated with calf alkaline phosphatase to prevent self-ligation. The template fragment was isolated after electrophoresis through a 0.8% agarose gel (Section 2.2.4.6) and ligated into the digested pCDNA3 by incubation with 1 µL T4 DNA ligase in T4 DNA ligase buffer in a total volume of 20 µL overnight at 16°C.

2.2.4.8 Ligation of specific gRNA sequences into pSpCas9n(BB)-2A-Puro (PX462)

Two separate gRNA plasmids were created to allow nicking of opposite strands of DNA either side of the mutated base in the APC gene. In each case the annealed oligos, which were

designed to have ends compatible with BpiI cut restriction sites, were ligated into the BpiI-linearized PX462 plasmid.

Complementary oligonucleotides encoding the two guide RNAs required for CRISPR-Cas9 editing of APC c.2557G>T, as detailed in Table 2.9, were annealed as follows. One µL forward oligonucleotide, 1 µL reverse oligonucleotide and 1 µL T4 DNA ligase were added to 7 µL H_2O, heated to 95°C for 5 minutes then allowed to cool for 2 hours at room temperature to anneal together.

The pSpCas9(BB)-2A-Puro plasmid (PX462) was digested with BpiI restriction enzyme for 2 hours at 37°C, treated with 1 µL calf alkaline phosphatase at 37°C for 1 hour to prevent religation of the plasmid, the enzyme denatured at 65°C for 10 minutes, and the plasmid cleaned using magnetic beads (2.2.3.6). Then 2 µL of digested plasmid were separately mixed with 1 µL of annealed oligos for each gRNA construct, 1 µL ligase buffer, and 0.5 µL T4 DNA ligase in a 10 µL reaction. They were incubated at 37°C for 1 hour, and then transformed into DH5α cells (Section 2.2.4.3) that were grown overnight on Luria-Bertani agar with 100µg/ml ampicillin. Single colonies were then grown in liquid media and plasmids were purified (Section 2.2.4.4).

2.2.4.9 Puromycin and geneticin sensitivity of HT29 cells

To determine the concentrations of puromycin and geneticin needed to select for HT29 cells containing the PX462 and template fragment plasmids, 24-well tissue culture plates were seeded with 5 x 10^4 HT29 cells/mL and grown in McCoy's 5A medium containing 10% fetal bovine serum (FBS) and penicillin/streptomycin (P/S) with increasing concentrations of the relevant drug (see Appendix 2). Four wells were seeded for each drug concentration with cell numbers in one well measured each day to allow growth to be assessed (Section 2.2.2.2). Three replicate plates were tested for each drug.

2.2.4.10 Transfection of plasmids into HT29 cells

HT29 cells were transfected using Lipofectamine reagents (Invitrogen) according to the manufacturer's protocols. Briefly, 7.5 x 10^5 HT29 cells were seeded in a 9.6 cm^2 well of a tissue culture plate and grown overnight. After 24 hours the growth medium was changed to

growth medium without antibiotics, and 3.75 µg of template fragment plasmid and each of the two PX462-gRNA Cas9 plasmids were added to 500 µL lipofectamine and 11.25 µL PLUS Reagent, mixed gently and incubated at room temperature for 10 minutes. Then 55 µL LTX reagent were added, mixed gently and incubated at room temperature for 30 minutes to allow lipofectamine-DNA complexes to form. Subsequently, 500 µL of this complex were pipetted onto the cells and incubated at 37°C in 5% CO_2 overnight. A control well of cells was treated with the transfection reagents without plasmids.

The following day, the cells were washed with PBS, lifted with TrypLE Express and resuspended to 50 mL in McCoy's medium containing 10% fetal bovine serum and penicillin/streptomycin. The cells were plated into three large 15 cm diameter tissue culture plates, topped up to 20 mL with growth medium, with a fourth plate used to grow the no plasmid control transfected cells. Following incubation at 37°C, in 5% CO_2 for 5 hours (to allow the cells to adhere), a further 20 mL medium containing 1 µg/mL puromycin and 3 mg/mL geneticin was added to each plate, and the cells were incubated for a further four days. The selective drugs, puromycin and geneticin, were removed by washing the cells with PBS and replacing the media with fresh drug-free media after approximately 90 hours of treatment.

Following the removal of selective drugs, the cells were incubated for a further 10 days before 96 individual colonies were each picked and transferred by pipette to a well of a 96-well tissue culture plate. The colonies were each incubated with 25 µL TrypLE Express for 10 minutes to disperse the cells, then 175 µL of medium were added and the plate was incubated at 37°C, in 5% CO_2 until the cells reached confluence. After washing, the cells were trypsinised and split into corresponding wells of three new 96-well plates and incubated for a further 2-3 days. Each plate was used for analysis of the clones, maintenance of the clones, or for freezing cells down for long-term storage.

For analyses, the cells for each clone in one of the three plates were washed, trypsinised and recovered into a microcentrifuge tube. After pelleting (5 minutes at 1300 x g), the medium was removed and the cells were washed in 10 mM Tris-Cl, pH8 before being resuspended in 10 µL 10 mM Tris-Cl and boiled for 15 min. Two microlitres of the cell suspension were then used in a PCR reaction to amplify a fragment encompassing the *c.2557G>T* mutation in the APC gene, using the primers used to generate the template fragment previously (Appendix 1). The PCR mixture and thermocycler conditions are detailed in Appendix 1. Following PCR, 10 µL of the PCR reaction were incubated with 0.5 µL XhoI at 37°C overnight. Digestion of the PCR

product into two fragments indicated successful incorporation of the template fragment into the genomic DNA of that clone.

2.2.4.11 Immunoprecipitation of the APC protein

Flasks of cells were washed and harvested by trypsinisation (2.2.2.1). The cells were pelleted by centrifugation at 3000 x *g* for 5 minutes and washed twice in ice-cold PBS before being resuspended in 2 mL of lysis buffer (Appendix 2) and sheared by passing through a 19-guage needle five times. The cell lysate was cleared by centrifuging at 13,000 x *g* for 20 minutes at 4°C and the protein concentration of the supernatant was measured by Bradford assay. Two aliquots, each containing 1 mg of protein in 1 mL of volume for each lysate, were used for immunoprecipitation, using two antibodies, raised against the N- (ab58) and C- (ab15270) terminals of the APC protein respectively.

Protein G-sepharose (GE Healthcare) was equilibrated by washing 500 µL of suspension three times in lysis buffer, before being resuspended in 1 mL lysis buffer to give a 50% suspension. For each immunoprecipitation, 40 µL of this suspension was added to 1 mL of cell lysate and rotated gently for 1 hour at 4°C, centrifuged at 3000 x *g* for 5 minutes, and the supernatant (cleared lysate) was transferred to a fresh microcentrifuge tube. One µg of antibody was added to the cleared lysate and incubated at 4°C for 1 hour. Then 40 µL of the 50% protein G-sepharose suspension was added and rotated at 4°C for 1 hour to capture the antibody and any attached immunoprecipitate to the sepharose. The lysate was centrifuged at 3000 x g at 4°C for 5 minutes and the supernatant discarded. The sepharose was washed four times with 1 mL lysis buffer, centrifuged and the supernatant discarded t each time. Following the final wash, the sepharose was resuspended in 100 µL of SDS-PAGE loading buffer, boiled for three minutes and proteins were resolved on an SDS-PAGE gel for Western blotting (Section 2.2.4.12).

2.2.4.12 Western blotting

Proteins were first resolved by polyacrylamide gel electrophoresis using 4-20% polyacrylamide gels (Table 2.11) in a Mini-PROTEAN Tetracell tank (BioRad) at 200V for 45 minutes, with top and bottom loading buffers as described in Appendix 2. Dual-coloured Precision Plus protein standards (BioRad) were run on each gel.

Separated proteins were transferred to polyvinylidene fluoride (PVDF) membrane (Table 2.11) using a Mini-Protean II transfer system (BioRad) with ice-cold transfer buffer (Appendix 2). Transfer was carried out at 100 V for 1 hour. A frozen plastic iceblock was added to the transfer tank directly from the freezer to prevent overheating during transfer. In addition, the transfer tank was placed on a magnetic stirrer with a magnetic flea ensuring continuous movement of the transfer buffer in the tank to maintain a cool temperature. Visualisation of the dual-coloured ladder on the PVDF membrane following transfer was used to confirm successful transfer of proteins.

Non-specific binding of antibodies was blocked by incubation of the PVDF membrane with 5% (w/v) non-fat milk in TBST (Appendix 2) for 1 hour at room temperature. After two five minute washes in TBST, the membrane was incubated overnight at 4°C in TBST with 2% (w/v) non-fat milk and primary antibody (directed against the N- or C-terminal of APC; both diluted 1:1000; Table 2.5). Three further 5 minute washes with TBST were followed by incubation with an appropriate secondary antibody (anti-mouse or rabbit, both diluted 1:2,500; Table 2.6) in TBST with 2% (w/v) non-fat milk for 90 minutes at room temperature. After three more 5 minute washes with TBST the membrane was exposed to chemiluminescent substrate (Table 2.11). Protein bands were visualized using a UVITEC gel documentation system and images captured with Nine Alliance software (UVITEC, Cambridge, UK).

2.2.5 *Bacteroides fragilis* toxin

2.2.5.1 Culturing *B. fragilis*

The enterotoxigenic *Bacteroides fragilis* (ETBF) strain 86-5443-2-2, originally isolated from a piglet with diarrhoeal disease and shown to produce the bft-2 toxin subtype (Franco et al., 1997) was used for these experiments. A non-toxigenic *B. fragilis* (NTBF) strain NTC9343 was used as a control. Both strains were kindly provided by Prof Cynthia Sears (Johns Hopkins University, Baltimore, USA).

Both strains were maintained on sheep blood agar plates under microaerophilic conditions, generated using a commercial Anaero-Pouch (Mitsubishi Chemical Gas Co, Japan) added to a gas jar (Oxoid, Basingstoke, UK). The plates were incubated at 37°C and sub-cultured twice a week.

To harvest the *B. fragilis* toxin (BFT), the bacteria were sub-cultured into 200 mL volumes of *B. fragilis* broth (Appendix 1; Franco, 1997) and incubated in gas jars for three days at 37°C under constant rotation at 120 rpm.

2.2.5.2 Concentrating the *B. fragilis* toxin (BFT)

The *B. fragilis* toxin (BFT) is released from the surface of toxigenic strains of these bacteria during growth in broth culture (Sears, 2001). Accordingly, the three day broth cultures were centrifuged (2330 x *g* for 10 minutes), the resultant supernatant was passed through a 0.22 µm syringe filter (to remove any residual bacteria) and either frozen at -80°C for future use or the proteins were precipitated at 4°C in ammonium sulphate. Briefly, 500 µL aliquots of supernatant were precipitated by the addition of a saturated solution of ammonium sulphate to give a dilution series of ammonium sulphate concentrations ranging from 0 to 60%. Proteins were pelleted by centrifugation at 4°C for 30 minutes at 21,000 x *g* then resuspended in 1/50th volume PBS. Each fraction was diluted ten-fold in medium and exposed to semi-confluent HT29 cells. The ETBF fraction that resulted in more than 50% of cells rounding was used for future experiments. The corresponding fraction from NTBF was used as a negative control.

Chapter 3 *CDH1* Gene Mutation in Early-Onset, Colorectal Signet-ring Cell Carcinoma

3.1 Introduction

Retrospective studies of colorectal cancer databases worldwide suggest that signet-ring cell cancers (SRCCs) account for around 1% of all primary colorectal cancers (Dozois et al., 2008; Hyngstrom et al., 2012; Inamura et al., 2015; Z. Liang et al., 2018; Nitsche et al., 2013; Wang et al., 2016; Wei et al., 2016). These tumours usually present at a later stage, have a worse grade and correspondingly poor prognosis (Hyngstrom et al., 2012). Colorectal SRCCs are defined as tumours exhibiting at least 50% of cells displaying signet-ring morphology (Bosman et al., 2010) although there is increasing evidence that even a small number of signet-ring cells can be detrimental to prognosis (Inamura et al., 2015; Sung et al., 2008; Tan et al., 2015).

The unique features of these cancers suggest a different pattern of genetic alterations to conventional CRC (Inamura et al., 2015; Kakar & Smyrk, 2005; Ogino et al., 2006; Wei et al., 2016) although a clear consensus on the molecular characteristics of colorectal SRCC is lacking. This is illustrated by conflicting evidence regarding the microsatellite instability status of these tumours (Nam et al., 2018; Ogino et al., 2006) that may, in part reflect the lack of published studies to date.

Large epidemiological studies suggest colorectal SRCCs can develop at any age in males and females, and throughout the colon and rectum (Hyngstrom et al., 2012; Tan et al., 2015). However, in young people with colorectal tumours with a signet-ring component, occurrence is more frequent in males (Wang et al., 2016) and the tumour site is more distal, particularly in the sigmoid colon and rectum (Chang et al., 2012; Wang et al., 2016).

There is mounting evidence that increasing numbers of young people presenting with early-onset colorectal cancer (EOCRC) are likely to have evidence of signet-ring histology (Chang et al., 2012; Tan et al., 2015), and that these patients have a worse prognosis than similar-aged patients with conventional adenocarcinoma, even with a low (<50%) signet-ring component to their tumour (Huang et al., 2016; Tan et al., 2015). Unlike conventional carcinomas, colorectal SRCCs produce large amounts of intracellular mucus that acts to push the nucleus to the edge of the cell resulting in the classical signet-ring appearance to the cell. One consequence of this altered morphology is a loss of cell-cell adhesion that, in turn, fits with the ability of colorectal

SRCC to diffusely infiltrate the stroma, leading to increased invasion and metastasis, and a poorer prognosis (H. C. Kim et al., 2002).

E-cadherin, encoded by the gene *CDH1*, is essential for maintaining intercellular stability in the intestinal epithelium via cell-cell adhesion and cell polarity. Accordingly, the loss of E-cadherin is considered to contribute to early stage carcinogenesis (Hirohashi, 1998). Reduced E-cadherin expression has been reported as a feature of colorectal CRC (Borger et al., 2007), and while signet-ring cells can show markedly reduced expression of this protein (H. C. Kim et al., 2002), this is not always the case (Chang et al., 2012; Wang et al., 2016).

Hereditary diffuse gastric cancer (HDGC) is typified by the presence of diffuse foci of signet-ring cancer cells (Charlton et al., 2004), with mutations in the *CDH1* gene identified in gastric (Guilford et al., 1998; Hakkaart et al., 2019; Norero et al., 2019) as well as in other malignancies (Xicola et al., 2019). Against this background, our colleague Prof. Frank Frizelle was intrigued to notice a large number of very young patients presenting with colorectal SRCC while working at the Aga Khan University Hospital in Karachi, Pakistan. In this chapter I describe the cohort of Pakistani colorectal SRCC patients diagnosed at or under 40 years of age, and investigate the possibility of mutation in the *CDH1* gene playing a causative, and possibly familial, role in for colorectal CRC in these young patients.

The objectives were to:

1. Determine the levels of E-cadherin expression in these colorectal SRCC samples.
2. Investigate the presence of *CDH1* mutations, and to correlate these with E-cadherin expression.
3. Establish the microsatellite stability status for each of the tumours.

3.2 Methods

Formalin-fixed, paraffin embedded (FFPE) tissues from 24 patients identified with colorectal SRCC at the Aga Khan University Hospital in Karachi, Pakistan were obtained, either as blocks or pre-cut sections. One sample that had no details with regards age and gender of the patient or location of the tumour was removed from the cohort. The study was granted ethics approval by the Ethics Review Committee, The Aga Khan University, Karachi, Pakistan.

One section from each tumour was stained with haematoxylin and eosin (H&E) for histological assessment by a pathologist (Section 2.2.1.2). Additional immunohistochemical staining was performed on sections of each sample for the mismatch repair proteins MSH6 and PMS2 in a 2-stain method for microsatellite instability testing (Hall et al., 2010). Additionally, staining for MLH1, another mismatch repair protein, was carried out for samples that were negative for PMS2 (Section 2.2.1.2). Another slide from each tumour was also stained for E-cadherin (Section 2.2.1.2).

DNA was extracted from up to three tissue sections (Section 2.2.1.1) and genomic DNA encompassing the promoter and all exons of the *CDH1* gene was amplified by a 2-step PCR protocol followed by sequencing by Illumina MiSeq. All molecular and downstream analysis protocols are detailed in Section 2.2.3.

3.3 Results

3.3.1 Histological characterisation of the Pakistani colorectal cancer cohort

Sections of tissue were received for nine of the 23 samples, while the remaining 14 samples arrived as FFPE blocks. Two and ten micron thick sections were cut from these blocks and used for staining of the tissues and extraction of DNA, respectively.

The hematoxylin and eosin (H&E) stained section of each tumour was reviewed by a pathologist according to WHO criteria for SRCC (Bosman et al., 2010). Three samples were diagnosed as mucinous adenocarcinomas (MA) rather than SRCC as they exhibited extracellular mucin and the cells lacked the classical signet-ring appearance. Although both MA and SRCC overproduce mucin, they exhibit different biological behaviours and survival outcomes (Hyngstrom et al., 2012) therefore these three MA samples were excluded from further analysis. Sixteen of the remaining 20 samples met the criteria of signet ring cell cancer based on the finding that >50% of the tumour cells had typical signet ring morphology (Barresi et al., 2017), while the remaining four were classified as CRC with a partial signet ring component on the basis of having fewer than 50% signet rings. Males were over-represented in the cohort (n=16), the median age was 27.5 years (range 8-40) and, while the tumours in our study were found at sites throughout the colon, 75% (15 of 20) were found in or below the sigmoid colon (Table 3.1). This predominantly distal distribution was not gender-specific.

Table 3.1. Clinical characteristics of the 20 SRCC samples received as part of the Pakistani colorectal cancer cohort.

	Total	
	N	%
Total	20	
Gender		
Male	16	80
…Female	4	20
Age at diagnosis (years)		
<21	6	30
21-30	6	30
31-40	8	40
Tumour site		
Proximal (right-sided)	5	25
Distal (left-sided)	15	75

3.3.2 DNA extraction and amplification.

DNA was extracted from the FFPE samples as described (Section 2.2.1.2) and the promoter and exons of *CDH1* were amplified according to the methods described in Section 2.2.1.1. Nine samples sent as pre-cut sections had to be excluded due to poor DNA yields and quality as evidenced by inconsistent amplification resulting in only a few amplicons of varying intensity being visible per sample (Figure 3.1). Of note, the few amplifiable regions of *CDH1* in the samples supplied as tissue sections were similar across samples, perhaps reflecting the dynamics of those PCR reactions or the relative degradation of DNA at those regions of *CDH1*. In contrast the DNA extracted from the tissues received as blocks was of sufficient quality (Figure 3.1). This may, in part, reflect the amount of material available for extraction on the individual slides. Whereas ten micron sections were cut from those tissues received from Pakistan as blocks, the remaining nine tissues were received as pre-cut sections and were therefore likely to be considerably thinner. Only three sections per patient were available for DNA extraction for these nine samples, which meant only six microns thickness of material per sample was available for extraction.

Accordingly, the number of samples available for analysis of the *CDH1* gene expression was reduced to 11. This group consisted of nine males and two females, with a median age of 28 years (range 15-40 years).

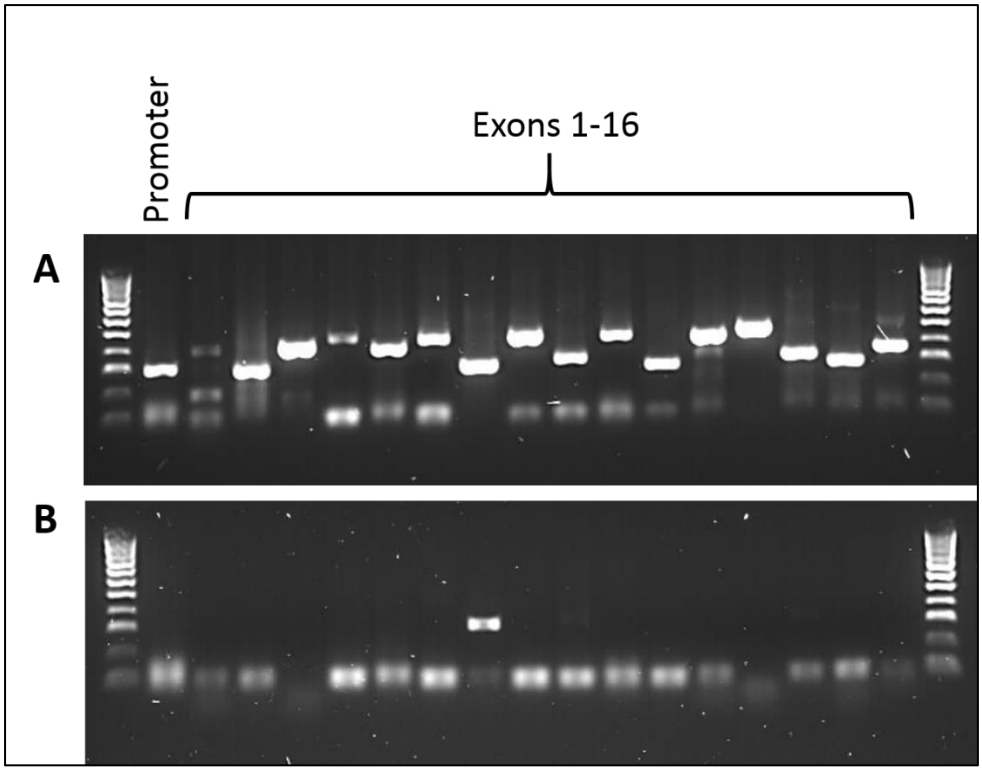

Figure 3.1 PCR amplification of *CDH1* promoter and exons. Amplicons from PCR reactions to amplify promoter region and exons 1-16 of *CDH1* gene from (A), DNA extracted from a sample supplied as a FFPE block and (B), DNA extracted from a sample supplied as cut sections.

3.3.3 Immunohistochemical staining for E-cadherin

The remaining 11 sections were immunostained to determine the presence of E-cadherin (Section 2.2.1.2), with staining of signet ring cells graded as negative, weak positive or positive (Wang et al., 2016). As with the H&E staining, the level of staining was scored by a pathologist blinded to the identity of the samples.

Immunohistochemistry revealed E-cadherin staining was absent (n=6) or notably weak (n=5) at the membranes of the signet-ring cells in the 11 tumours when compared to adenocarcinoma cells or normal tissue within the same samples (Figure 3.2; Table 3.2).

Table 3.2 E-cadherin staining by signet ring cell component of the 11 SRCC samples included in the final analysis.

Histology	Age	Gender	Location	E-cadherin staining
Colorectal SRCC (>50% signet ring cells)	15	F	anal growth	negative
	24	M	rectosigmoid	negative
	24	M	anal canal	weak positive
	28	M	rectal	negative
	37	M	rectosigmoid	weak positive
	39	M	rectum	weak positive
	21	F	gut nodule	weak positive
	35	M	hepatic flexure	negative
	40	M	cecum	negative
CRC (with <50% signet ring cell component)	20	M	cecum	weak positive
	30	M	anus	weak positive

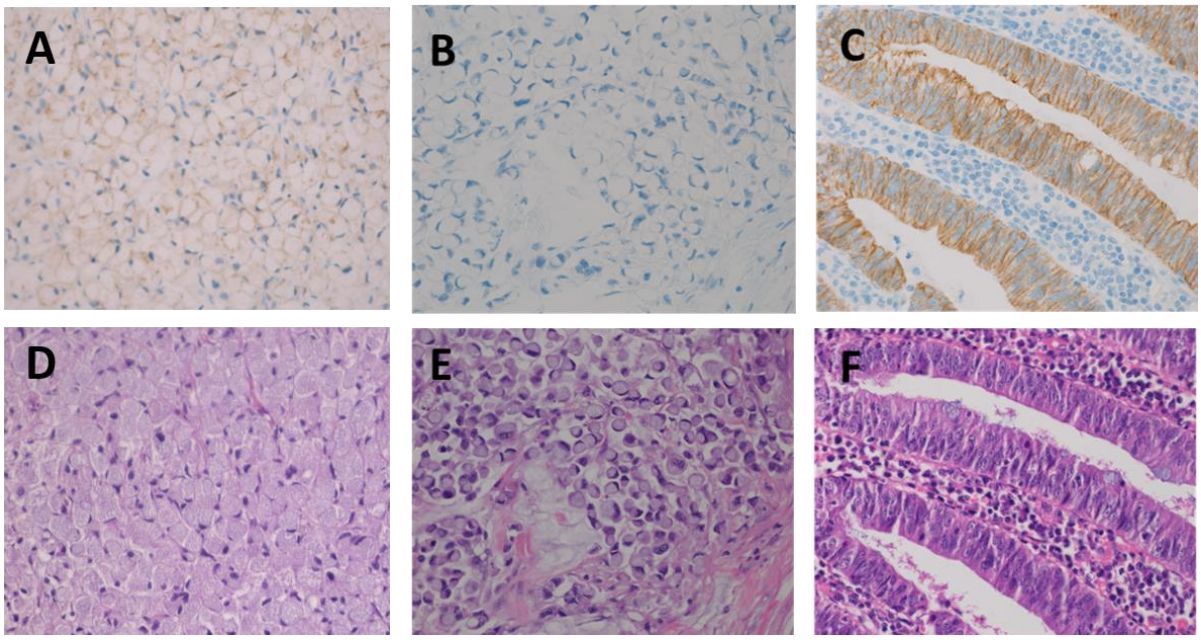

Figure 3.2 E-cadherin staining of signet ring cell cancer. Representative immunohistochemistry with (A) weakly staining positive and (B) negative E-cadherin staining of signet ring cells, and (C) adenocarcinoma cells showing typical E-cadherin staining, alongside sequential H&E staining of the same sections (D, E & F respectively).

3.3.4 Molecular analysis of the *CDH1* gene

The *CDH1* gene sequence was determined by running pooled, indexed PCR products on an Illumina MiSeq instrument following the protocol described by Hakkaart *et al*. (Hakkaart et al., 2019). This approach involved the generation of dual-indexed amplicon sequencing libraries for each of the 11 samples with amplifiable DNA, that were diagnosed as either colorectal SRCC or CRC with signet-ring cell component, using a two-step PCR strategy (Section 2.2.3).

Briefly, 17 target sites (16 exons and the promoter region) were amplified in separate PCR reactions and checked by electrophoresis on a 1.5% agarose gel (Figure 3.1). Following successful amplification, the PCR products for each sample were pooled and further amplified using unique pairs of primers containing indices that would allow for the identification of each sample within an Illumina MiSeq sequencing run.

Raw paired end reads were cleaned with Trimmomatic v.0.35.25. Cleaned reads were aligned to the human reference genome (GRCh37/hg19) using the Burrows–Wheeler Aligner v.0.7.10 (H. Li & Durbin, 2009). Single nucleotide variants and insertion/deletion variants were called using 'The Genome Analysis Toolkit' (GATK) v.3.6 (McKenna et al., 2010). The effects of variants were predicted using SnpEff v.4.2 (Cingolani et al., 2012). The trimming of reads, sequence alignments, variant calling and variant effect prediction were performed by Dr Christopher Hakkaart using a bioinformatics pipeline developed by him. All putative mutations were visually inspected using the Integrative Genomics Viewer (Robinson et al., 2011). The functional consequences of missense variants were predicted *in silico* using five different models that included SIFT (Kumar et al., 2009), Provean (Choi et al., 2012), Mutation Assessor (Reva et al., 2011) and PolyPhen2 (using both HumDiv and HumVar datasets) (Adzhubei et al., 2013). Mutations predicted to be deleterious by at least 3/5 predictive models were considered of interest.

Raw sequencing data were cleaned and aligned to the reference genome (GRCh37/hg19) as explained in Section 2.2.3.11. Variants were called and their effects predicted as detailed in the same section, with all variants visually inspected using the Integrative Genomics Viewer (Robinson et al., 2011). Multiple *in silico* models were used to predict the pathogenicity of mutations with a mutation being considered of interest if at least 3/5 models predicted it to be pathogenic (Section 2.2.3.11).

This analysis of the data generated using next generation sequencing of the *CDH1* gene revealed eight mutations meeting our inclusion criteria regarding absolute (>10) and relative (>20%) read depth in 7 of the 11 (64%) samples that yielded DNA of sufficient quality for analysis (Table 3.3). These samples included five signet-ring cancers and two CRCs with a partial signet-ring component.

Mutation status was not associated with the histological classification of the cancer. In line with the spread of reported mutations in the COSMIC (Catalogue of Somatic Mutations In Cancer) database, the *CDH1* mutations were spread throughout the gene with no clear hotspots (Figure 3.3), although three out of the eight mutations occurred in the third ectodomain. All mutations were classified as variants of unknown significance. Six mutations had not previously been described. Allele counts indicated that five mutations found in four of the seven samples were possibly germline (Table 3.3) where the mutant allele accounted for 50% or more of the reads.

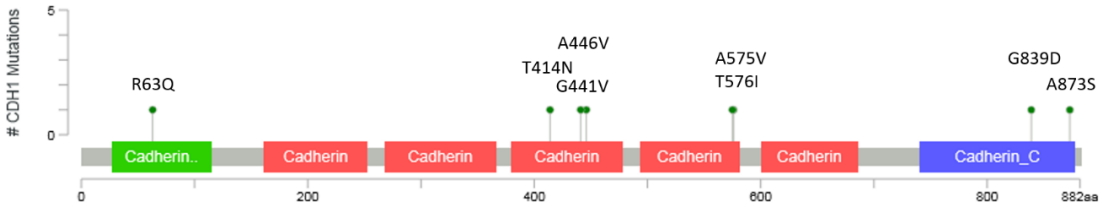

Figure 3.3 Distribution of *CDH1* mutations within the SRCC cohort. The locations of the mutations found in this study are indicated by lollipops along the length of the CDH1 protein. The mutations occurred in multiple different domains throughout the protein: the pre-protein (green), ectodomains (red) and the intracellular domain (blue). The plot was generated using Mutation Mapper (www.cbioportal.org) and manually curated.

Table 3.3 Characteristics of *CDH1* mutations.

Domain	Mutation		Allele Count[1]	Mutation Assessment[2]	E-cadherin staining	PMS2 staining	Histology
	DNA	Protein					
Germline							
Cytoplasm	c.2516G>A	p.G839D	2,35	damaging	negative	positive	SRCC
EC4	c.1724C>T	p.A575V	25,25	damaging	weak positive	positive	SRCC
EC4	c.1727C>T	p.T576I	25,25	damaging			
EC3	c.1241C>A	p.T414N	56,64	benign	negative	positive	SRCC
EC3	c.1337C>T	p.A446V	9,36	benign	weak positive	negative	mixed
Somatic							
Cytoplasm	c.2617G>T	p.A873S	197,72	damaging	negative	positive	SRCC
EC3	c.1322G>T	p.G441V	42,14	damaging	weak positive	positive	SRCC
Preprotein	c.188G>A	p.R63Q	37,18	benign	weak positive	positive	mixed

[1] (consensus allele, mutant allele), where mutant alleles present in at least 50% of reads were considered to be germline while those present in less than 40% of reads were determined to be somatic. [2] considered damaging if 3/5 or more software tools predicted the mutation to be possibly or probably damaging. Nonsense or frameshift mutations were all considered damaging without predictive tools.

3.3.5 Clinical characteristics of samples with predicted germline mutations

Of the four samples with putative germline mutations, three were from patients diagnosed with colorectal SRCCs with the other case being a CRC with partial signet-ring morphology. Moreover, in silico analysis of the *CDH1* gene in two of these tumours suggested the mutations are damaging (Table 3.3).

Sequence analysis revealed that one SRCC tumour had a germline mutation (p.G839D) that was universally predicted by all *in silico* tools to have a deleterious effect on the E-cadherin protein. Additionally, nearly all the reads showed the mutated form, suggesting the mutation may be inherited. While mutations at codon 839 have not previously been described, all substitutions in the COSMIC database within ten codons either side of codon 839 are predicted to be pathogenic, likely due to the location within the intracellular catenin-binding domain (Figure 3.3). Disruption of catenin binding to the intracellular domain of E-cadherin would

likely destabilise the cellular architecture via loss of binding of the catenin complex to actin filaments (Mege & Ishiyama, 2017). There was no evidence E-cadherin staining of signet ring cells in this tumour.

The other two possible germline deleterious *CDH1* mutations occurred in the same sample. Two adjacent amino acids (p.A575V, p.T576I) were mutated on the same allele, in an area encoding an extracellular domain (Table 3.3, Figure 3.3). *In silico*, these adjacent mutations would be considered likely to affect the ability of the mutant E-cadherin to form effective junctions with adjacent cells. Weak E-cadherin staining of signet-ring cells in this sample may equate with this predicted loss of functional activity.

Sequence analysis of the *CDH1* gene in a further SRCC sample suggested evidence of a p.T414N substitution. *In silico*, this mutation was predicted to be benign, as was the p.A446V mutation found in a CRC with signet-ring cell component (Table 3.3). Signet-ring cells in the sample with the p.A446V mutation were weakly positive for E-cadherin staining. In contrast, the sample with the p.T414N mutation did not stain for E-cadherin. The p.T414N mutation has previously been identified and designated as of unknown significance and, currently, there is no evidence to suggest that or the p.A446V mutations have clinical significance. If not, the loss of E-cadherin in the sample with the p.T414N mutation may be due to other mechanisms, such as epigenetic modification (Darwanto et al., 2003).

3.3.6 Clinical characteristics of samples with predicted somatic mutations

The remaining *CDH1* mutations were considered somatic, given the low relative allele counts (<40%; Table 3.3). This discovery of multiple somatic mutations is not surprising given the late stage at diagnosis of these tumours where there is increased likelihood of a high and heterogeneous mutational load of somatic mutations.

One of the somatic mutations, p.R63Q, has been described previously as benign (https://www.ncbi.nlm.nih.gov/snp/rs587780117). Codon 63 is located within the pre-protein so is not part of the mature active protein. Therefore, providing the mutation does not affect the intracellular targeting of the protein, loss of E-cadherin staining in this sample may be due to another mechanism. However, in the absence of evidence of another mechanism R63Q cannot be ruled out as the cause of the loss of E-cadherin staining in this sample.

3.3.7 Microsatellite instability was not a common feature in this cohort

Microsatellite instability (MSI) is found in 15-20% of sporadic colorectal cancer and considered to be a positive prognostic factor (F. A. Sinicrope et al., 2006). MSI is also reported in colorectal SRCC (Ogino et al., 2006; Rosty et al., 2014) although this is not universal (Nam et al., 2018) and, when present, may not have a role in cancer-related mortality (Inamura et al., 2015). Accordingly, tissue sections were immunostained to identify microsatellite markers MSH6, PMS2 and, in the event of negative PMS2, MLH1 to better understand the potential contribution of MSI to SRCC prognosis in this cohort. Again, the stained sections were scored by a pathologist blinded to the identity of the samples.

While it is standard practice in New Zealand diagnostic laboratories to stain for all four MMR proteins (MSH2, MSH6, MLH1 and PMS2), just MSH6 and PMS2 were stained in the first instance. This was for two reasons, namely cost and to maximize the number of slides available for DNA extraction, particularly for those samples that were only supplied as a limited number of pre-cut sections, and was on the advice of an anatomical pathologist although it has also been reported elsewhere (Hall et al., 2010). This approach is based on the observation that the MSH2 and MSH6 form a heterodimers during mismatch repair while PMS2 and MLH1 do likewise. While MSH2 can also partner with MSH3, MSH6 can only pair with MSH2 so lack of MSH2 staining would also lead to lack of MSH6 staining therefore MSH6 can be used as a proxy for MSH2. Similarly PMS2 can be used as a proxy for MLH1. Samples deficient in PMS2 or MSH6 staining would then have sections stained with MLH1 or MSH2 respectively, although all samples were MSH6 positive so no MSH2 staining was ultimately required for this study. This protocol is not, however, a full-proof method of assessing MMR deficiency as in a very small number of cases, MMR deficiency can be present despite positive MSH6 and PMS2 staining (Pearlman et al., 2018). However, due to the considerations mentioned above and the relative rarity of the 2-stain method failing to identify MMR deficiency it was decided to proceed with the 2-stain method.

All of the 11 samples included in the sequence analysis above were positive for MSH6 staining, however two samples were negative for PMS2 (Figure 3.4), loss of which is associated with colorectal cancer. Interestingly, both samples (one SRCC and one CRC with signet-ring cell component) that stained negative for PMS2 showed staining for E-cadherin, albeit weak, suggesting perhaps that microsatellite instability and loss of E-cadherin represent different pathways of colorectal signet-ring cell development. This argument is further strengthened by

noting that one of these two samples had no *CDH1* mutation while the other sample had a mutation that was predicted to be benign. Subsequent testing of these two samples for MLH1 staining showed one sample lacking MLH1 while the other sample was inconclusive due to non-reactivity of the control tissue added to each slide holding a tumour section prior to staining.

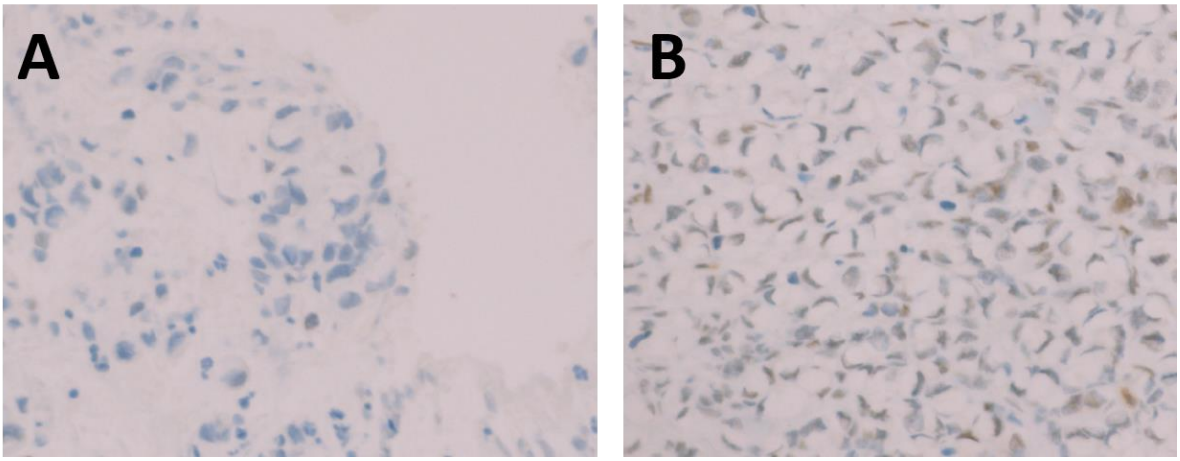

Figure 3.4. PMS2 staining of signet-ring cells. Immunohistochemistry showing (A) a sample with PMS2 negative signet-ring cells and (B) a contrasting sample with nuclear staining for PMS2 in signet-ring cells.

3.4 Discussion

While the prevalence of colorectal SRCC in Pakistan is currently unknown, a growing number of studies show high incidences of early-onset CRC across the Asian continent (Chew et al., 2009; Gupta et al., 2010; Laskar et al., 2014; Wang et al., 2016) including Pakistan where younger age and advanced stage of disease at diagnosis were identified in two unrelated studies (Amini et al., 2013; Bhurgri et al., 2011). While neither of the Pakistani studies specifically reported the percentage of young patients with SRCC, early-onset CRCs are more likely to demonstrate signet ring cell differentiation when compared to CRCs seen in patients aged over 40 years (Chang et al., 2012; Q. Li et al., 2014). Collectively, these studies support anecdotal evidence of high rates of colorectal SRCC in young Pakistani people. Against that background, our colleague's observance of high rates of colorectal SRCC in young Pakistani patients at a hospital in Karachi, provided us with an intriguing cohort in which to study this disease.

Studies suggest the younger the age of the patient, the poorer the outcome (Huang et al., 2016; Poles et al., 2016). The youngest patient in the cohort was only eight years old and 35% (7 of

20) of the patients diagnosed with CRC with a signet ring cell component (irrespective of whether signet ring cells were a major (>50%) or minor (<50%) component of the tumour) were younger than 22 years of age, a cut-off which has been used to define CRC patient populations as paediatric (≤21 years) and early-onset adult (22-50 years), respectively (Poles et al., 2016). No family history was available for any patients in our study, but the early age of onset suggests that some of these cancers may be familial rather than sporadic (Wei et al., 2016). E-cadherin loss is a well-documented feature of familial SRCC gastric cancer (Guilford et al., 2007), and while colorectal SRCC is far less studied, the loss of this adherens junction protein has been reported as a feature of this disease too (H. C. Kim et al., 2002), and may explain the increased invasiveness and metastatic potential of colorectal SRCCs.

Eleven samples in our cohort of patients with colorectal SRCC had DNA of sufficiently good quality for molecular analysis by next generation sequencing. Using this technique, we found evidence of one or more mutations in the *CDH1* gene in seven (64%) samples that included five predicted germline and three predicted somatic mutations. Unfortunately, neither matched normal tissue nor peripheral blood samples were available to confirm the germline status of the mutations. The use of allele counts to predict the germline status of is a crude and inaccurate method which may be affected by aneuploidy within tumour cells and the presence of non-tumour cells. As such the assignment of possible germline status to any samples in this chapter is made due to the absence of normal tissue and is only speculative.

The absence of any common mutation between patients suggests that if there are familial cases in our cohort it is not due to a single founder mutation within that population. It is possible however that multiple mutations may arise in a population, as evidenced by the discovery of multiple familial *CDH1* mutations in New Zealand Māori (Hakkaart et al., 2019; Humar et al., 2009).

It has been postulated that the occurrence of multiple germline *CDH1* mutations in a population may reflect a selective advantage in these populations. One explanation revolves around *Listeria monocytogenes*, a food-borne pathogen causing gastroenteritis and, in serious cases, meningitis or miscarriage in pregnant women (da Silva Tatley et al., 2003). This bacterium internalizes into epithelial cells via the binding of the bacterial internalin-A protein to the N-terminal, extracellular domain of E-cadherin (Hamon et al., 2006). Potentially, mutant E-cadherin protein may interfere with the ability of *L. monocytogenes* to bind and internalise

in epithelial cells. However, whether or not there is (or was) a selective advantage conferred by these mutations remains to be determined.

The low mutant allele count for some of the mutations may indicate that for some samples mutation of *CDH1* is a late event in the development of colorectal SRCC. However, the two SRCC cases with seemingly deleterious germline mutations and others with high mutant allele counts suggest that E-cadherin loss is an important early step in a subset of SRCC. This is highlighted by our finding that all three paediatric patients (ranging in age from 15 to 21 y) in this group had evidence of mutations in their respective *CDH1* genes. One tumour contained a predicted germline mutation, despite the signet ring cells staining weakly positive for E-cadherin. The two other paediatric patients had tumours that were classified as colorectal SRCCs (>50% signet rings), each with evidence of predicted damaging somatic rather than germline mutations in the *CDH1* gene, and again the signet ring cells in one of these tumours showed weak evidence of E-cadherin expression. Collectively, these findings suggest a lack of concordance between the molecular and phenotypic expression of E-cadherin although it is possible that weak E-cadherin immunoreactivity may not necessarily equate with functional activity (Becker et al., 1994).

Microsatellite instability (MSI) has previously been associated with colorectal SRCC (Ogino et al., 2006; Rosty et al., 2014) but we found low levels (2/11, 18%) in our cohort. Furthermore, one of the two was identified as a mixed signet ring/mucinous adenocarcinoma sample, meaning the prevalence of MSI is even lower, with only one out of the seven samples (14%) classified as SRCC having microsatellite instability. It is to be noted, however, the evidence for MSI in colorectal SRCC is based on small scale studies and case reports and is not a universal conclusion in all studies. A recent study of five Korean SRCC patients found all cases were microsatellite stable (Nam et al., 2018).

Microsatellite instability in sporadic colorectal cancer tends to be associated with proximal tumours (H. Kim et al., 1994; Thibodeau et al., 1993) and occurs in females more than males (F. Sinicrope et al., 2010), neither characteristic seen in our (admittedly small) cohort of SRCC patients. Additionally, MSI in sporadic CRC has a better prognosis than microsatellite stable (MSS) cancers, while colorectal SRCC has a very poor prognosis irrespective of MSI status. In light of these observations it may not be surprising that we find low levels of MSI in our young, predominantly male cohort with mostly distal tumours. Both our cohort and the MSS cohort in the Korean study of SRCC described above (Nam et al., 2018) had low median ages

(28y and 40y respectively) therefore it may be that our findings reflect a shift from lower to higher levels of MSI with increasing age and that MSI may not be a strong feature of early-onset colorectal SRCC.

A major limitation of this study was that clinical data for the cohort included only age, gender and tumour site. No other clinical variables were available and there was no follow-up or family history for any of the samples. The lack of this information meant that no assessment of outcome in relation to E-cadherin expression (Wang et al., 2016) and/or *CDH1* mutation was possible. While limited follow-up was only ever going to be possible given the diagnoses ranged from 2015-2017, the poor prognosis associated with colorectal SRCC suggest that some useful conclusions may have been made had this information been available.

Similarly, the lack of family history prevents any evaluation of genetic predisposition to SRCC. This would also be informed by access to matched normal tissue or peripheral blood samples that could have confirmed the germline status of any mutations found. As well as limiting this study, the inability to show germline mutation of *CDH1* prevents any benefit to patient families that might be afforded by early screening or vigilance.

Additionally, the poor quality of the samples limited the scope of the study in several ways. We were only able to report on 14 of the original 24 samples due to the poor quality of DNA recovered from tissue sections. This limited the power of the study, particularly with the small number of starting samples. Due to the rarity of colorectal SRCC, many of the existing studies have very small cohorts. Hence, a study with 24 SRCC patients had the potential to significantly increase current knowledge around colorectal SRCC. However, notwithstanding the smaller sample number, we still regard this study as a useful addition to the currently sparse data on this rare condition.

In conclusion, we found E-cadherin expression to be absent (or at best very weak) in colorectal signet-ring cell carcinomas and that this corresponded in a small number of cases with germline mutation of the *CDH1* gene. For these patients, testing of family members for germline mutation may allow screening and early detection of tumours. Other colorectal SRCC cases showed low levels of somatic *CDH1* mutations. Collectively, these findings suggest that while loss of E-cadherin is a common feature of colorectal SRCC and colorectal cancers with a signet-ring cell component, mutation of the *CDH1* gene is not an underlying cause in the development of all cases.

Chapter 4 *CDH1* and *APC* Mutation in Early-Onset Colorectal Cancer

4.1 Introduction

Our finding, described in Chapter 3, that absent or weak E-cadherin expression coincided with multiple *CDH1* mutations within the cohort of Pakistani SRCC samples, was intriguing. However our ability to further assess the implications of this within that cohort was limited by sample quality and a lack of access to further samples (such as matched normal tissue) and/or clinical data, including patient outcomes and family histories. Our inability to confirm the germline status of those mutations with high allele counts, due to lack of matched non-tumour tissue, prevented us from drawing any conclusions as to whether screening of family members may have been of benefit. In addition, the lack of any family history of affected individuals limits the insights that can be made from our findings.

In light of these limitations we sought to investigate E-cadherin expression and *CDH1* mutation in a local cohort of early-onset colorectal cancer (EOCRC) patients where clinical information would be more readily available and sample quality more consistent. In addition we also sought to determine the presence of mutations in the *APC* gene for two reasons. Firstly, *APC* mutations occur in 60-80% of sporadic CRCs (Rowan et al., 2000) and mutations in this gene have long been established as an early event in colorectal tumourigenesis (Powell et al., 1992). Secondly we were aware that the number of SRCC cases in our cohort of EOCRC patients was likely to be very low (Plunkett et al., 2014).

Interestingly, there are studies that report the rate of *APC* mutation in early-onset CRCs to be lower than the rate seen in older CRC patients (Kothari et al., 2016), although this is not always the case (Willauer et al., 2019). While this may reflect different pathways to tumourigenesis in these individuals, another possibility is that these differences may reflect the use of different methodologies to look for mutations in a gene that has a large coding sequence spanning multiple exons over more than 100 kilobases of genomic DNA (Luchtenborg et al., 2004).

There have been many different approaches in assessing *APC* mutations, with different regions of *APC* investigated in each approach (Kothari et al., 2016; Luchtenborg et al., 2004). While there is a defined mutation cluster region within the final exon of the gene, there are many mutations occurring outside this region. Therefore, strategies that limit themselves to the mutation cluster region, or more widely to the final exon, will miss a proportion of mutations

found elsewhere in the gene. To address this we looked at regions of the *APC* gene encompassing the vast majority of mutations found in the COSMIC database. It is reasonably common for a tumour to have both *APC* alleles mutated. When this occurs there tends to be a 3' mutation in or near the mutation cluster region, plus a mutation further upstream on the other allele (Christie et al., 2013). Truncated APC proteins have different functionalities and so this distribution of mutations within tumour cells is not a random coincidence. Therefore screening for mutations upstream of the mutation cluster region is warranted and necessary for a fuller picture of APC-related colon carcinogenesis.

Frozen tumours and off-tumour tissue samples, obtained from 25 patients identified with early-onset CRC, were collected at Christchurch Hospital, New Zealand, between August 2005 and February 2019. None of the patients had treatment prior to surgery and known Hereditary Non-Polyposis Colorectal Cancer (HNPCC, Lynch syndrome) or Familial Adenomatous Polyposis patients were not included.

The objectives of this study were to:

1. Investigate the presence of *CDH1* and *APC* mutations in a cohort of early-onset CRC cases from New Zealand by next generation sequencing protocols.
2. Determine by immunohistochemistry the level of E-cadherin expression in tumours from the early-onset CRC patients and compare this to the DNA sequence information.
3. Use single nucleotide polymorphisms within the sequence obtained in objective 1 to assess any loss of heterozygosity within the *CDH1* and *APC* loci.

4.2 Methods

Samples from 25 patients with early-onset colorectal cancer, diagnosed at or under the age of 50, were obtained from the Cancer Society Tissue Bank New Zealand. The samples consisted of frozen tumour tissue as well as matched normal tissue taken at the same time of surgery. The study was granted ethics approval by the University of Otago Human Ethics Committee (H18/143).

DNA was extracted from 25 mg of frozen tissue (Section 2.2.1.1) and regions of DNA encompassing the promoter and all exons of the *CDH1* gene were amplified by a 2-step PCR protocol followed by sequencing on an Illumina MiSeq platform in the Department of

Biochemistry, University of Otago, Dunedin. All molecular and downstream analysis protocols are detailed in Section 2.2.3.

FFPE tissue blocks were obtained for each sample, sections cut and one slide for each case was stained with haematoxylin (to reveal nuclei) and eosin (to stain the extracellular matrix and cytoplasm) (H&E) for histological assessment by a pathologist (Section 2.2.1.2). In addition, a slide from each tumour was immunostained for E-cadherin (Section 2.2.1.2).

The microsatellite instability status of the samples, as determined by the expression of four mismatch repair genes, MSH2, MSH6, PMS2 and MLH1, was carried out as part of the clinical assessment of the tumours, using the same method detailed in Chapter 2 (section 2.2.1.2). This data was obtained from the clinical records.

4.3 Results

4.3.1 Characterisation of the cohort

The cohort of 25 early-onset CRC cases consisted of 14 (56%) women and 11 (44%) men aged 50 years or under at diagnosis (median 44 years, range 28-50 years) (Table 4.1). The tumours were predominantly distal and all but two were classed as adenocarcinomas by H&E staining, with an adenosquamous carcinoma, where tumours contain cells originating from both mucus-producing glandular cells and flat squamous cells that line epithelia, and a signet-ring cell carcinoma completing the cohort (Table 4.1). While the tumours were predominantly distal this is likely an under-representation of the proportion of distal tumours in early-onset CRC. Due to protocols used by the tissue bank, following removal of the colon, the rectum is not opened up prior to formalin fixation in order to preserve the resection margins. Therefore it is not possible for the tissue bank to obtain fresh rectal tumour tissue. In addition, neoadjuvant therapy on rectal tumour patients prior to surgery would have excluded inclusion of many of the rectal samples from our cohort (if they were accessible).

Table 4.1: Clinical features of the early-onset CRC cohort.

	Number	%
Total	25	
Gender		
Male	11	44
Female	14	56
Age at diagnosis (years)		
21-30	1	4
31-40	6	24
41-50	18	72
Ethnicity		
European	24	96
Filipino	1	4
Stage at diagnosis		
I	9	36
II	9	36
III	6	24
IV	1	4
Tumour site		
Caecum	1	4
Ascending colon	6	24
Hepatic flexure	1	4
Splenic flexure	1	4
Descending colon	3	12
Sigmoid colon	12	48
Rectosigmoid	1	4
Pathology		
Adenocarcinoma	23	92
Adenosquamous carcinoma	1	4
Signet-ring cell carcinoma	1	4

4.3.2 Immunohistochemical staining for E-cadherin

Sections of tumour tissue that were cut from FFPE blocks were immunostained for the E-cadherin protein (Section 2.2.1.2) and were independently assessed by a pathologist. As with the Pakistani cohort (Chapter 3), the level of staining of the tumour cells was graded as negative, weak positive or positive.

While frozen tissue was available for all 25 patients, FFPE blocks were not available for one patient, meaning that tumour tissue could not be stained for E-cadherin in this case. Of the remaining 24 samples only one tumour was classified as negative for E-cadherin, while a second sample stained weakly. Consistent with previous reports of E-cadherin status in

colorectal SRCC (Aitchison et al., 2020; Borger et al., 2007), the sample with negative E-cadherin staining was the sole SRCC sample in the cohort.

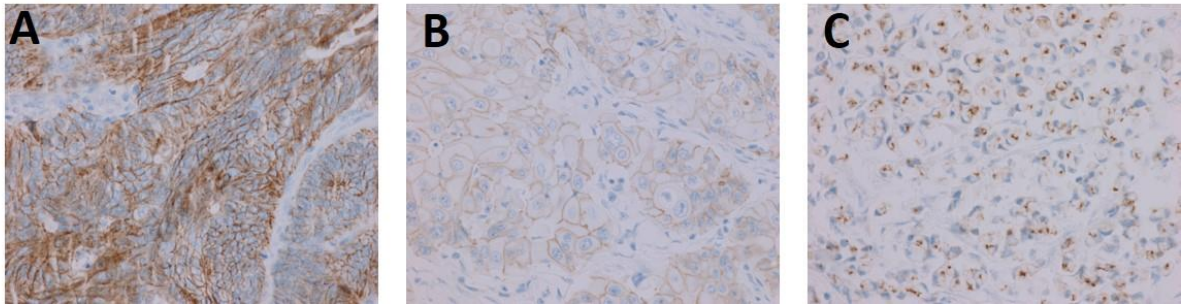

Figure 4.1. Representative immunohistochemistry for E-cadherin in EOCRC samples. A, positively-staining adenocarcinoma sample showing strong membrane-associated E-cadherin staining (brown) in tumour cells. B, an adenosquamous carcinoma sample showing weak E-cadherin staining at tumour cell membranes although the staining was consistent along the cell membrane. C, signet-ring cell carcinoma cells with small foci of E-cadherin at cell junctions within predominantly negatively staining cells showing classic signet-ring morphology. A pathologist classified this tumour as negative for E-cadherin due to the absence of E-cadherin along the membranes. Nuclei are stained blue.

The sample with weak E-cadherin staining was the sole adenosquamous carcinoma, whereas all the 22 other tumours that stained positive for E-cadherin were adenocarcinomas. Interestingly, within the adenosquamous carcinoma it was the squamous cells that stained very weakly, whereas stronger staining was seen in the glandular cells. This suggests that the squamous proportion of the carcinoma contributed to the weakness of the E-cadherin signal in this sample.

4.3.3 Molecular analysis of the *CDH1* gene

DNA was extracted from 25 mg of frozen tumour and matched normal tissue using a Qiagen DNeasy Blood & Tissue kit (Section 2.2.1.1). The *CDH1* gene sequence was determined by running pooled, indexed PCR products on an Illumina MiSeq instrument following the protocol described in Section 2.2.3. Again, mutations predicted to be deleterious by at least 3/5 predictive models were considered of interest while mutations with low absolute (<10) or relative (<20%) allele counts were excluded.

This analysis of the data revealed a single tumour sample from a 34 year old male that met the inclusion criteria. This sample yielded a *c.88C>A* variant in the *CDH1* gene, a missense variant resulting in a p.P30T substitution. The proline residue encoded by codon 30 of E-cadherin is moderately conserved between species and is within a very conserved region of the E-cadherin sequence, between codons 28 and 32 (Figure 4.2). The *c.88C>A* variant is seen at low frequency in databases such as 1000Genomes (MAF 0.00040) and ExAc (MAF 0.00094), and multiple predictive tools suggest this variant to be deleterious. In contrast, ClinVar considers *c.88C>A* as likely benign (https://www.ncbi.nlm.nih.gov/clinvar/variation/127933/).

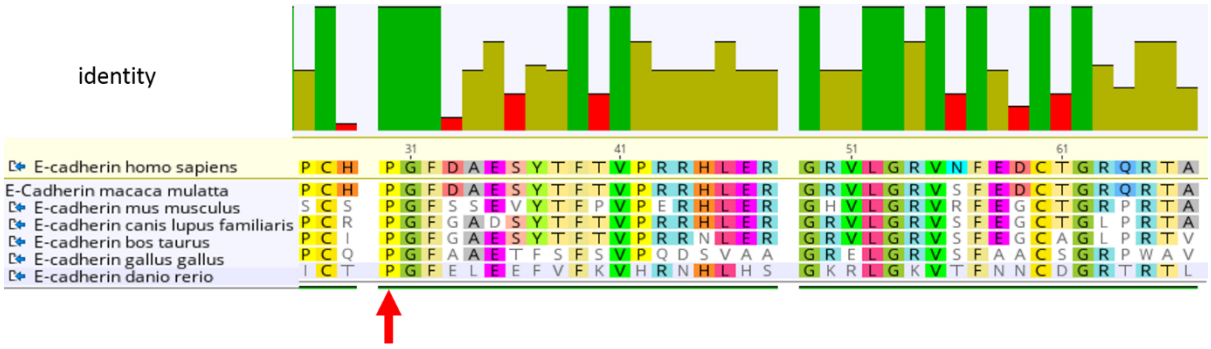

Figure 4.2. Conservation of the proline residue at codon 30 of E-cadherin. A section of the E-cadherin containing protein including the proline residue at codon 30 (red arrow) which is conserved to zebrafish (*Danio rerio*) and is adjacent to the strongly conserved glycine and phenylalanine residues at codons 31 and 32. Alignments created using Clustal Omega in Geneious Prime 2020.0.5 (Biomatters Ltd, Auckland, NZ).

Interestingly the p.P30T substitution created by the *c.88C>A* variant is located in the pre-protein, and therefore does not impact on the mature E-cadherin protein sequence. However it may play a role in the localisation of the protein to the cell membrane (Vogelaar et al., 2013) and, while proline and threonine are physiochemically similar, this change may subtly impact this potential targeting role.

Sanger sequencing of tumour tissue confirmed the presence of the variant, and sequencing of the matched normal tissue confirmed the germline status of the variant (Figure 4.3).

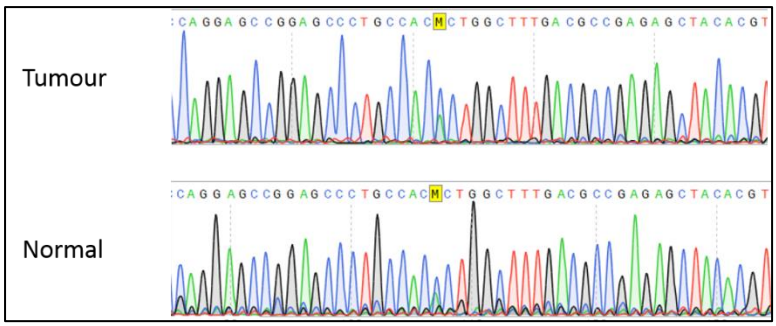

Figure 4.3. Confirmation of germline status of CDH1 c.88C>A mutation. Sanger sequencing of the *CDH1 c.88C>A* (recognised as a heterozygous base, M, in the sequence) mutation, confirming its presence in both tumour and matched normal tissue.

4.3.4 Molecular analysis of the *APC* gene

As detailed in the introduction to this chapter, there are many different approaches to assessing *APC* mutations that can, in part, relate to the region of the gene in which they are located. Accordingly, an approach that is limited to investigating mutations within the defined mutation cluster region within the final exon of the gene is likely to miss a proportion of mutations found elsewhere in the gene. To address this we looked at regions of *APC* encompassing the vast majority of mutations found in the COSMIC database.

Fifteen amplicons covering the *APC* sequences most commonly mutated in colorectal cancer were amplified for each of the 25 EOCRC samples, as detailed in Section 2.2.3 (Figure 4.4). A second round of PCR was performed using indexed primers as for the *CDH1* sequencing detailed above. Pooled products for each tumour and matched normal tissue samples were then sequenced (Sections 2.2.3.9 to 2.2.3.11).

Twenty four mutations meeting the inclusion criteria for absolute (>10) and relative (>20%) read number were found in 18/25 tumour samples. The mutations consisted of 14 nonsense mutations, 6 frameshift mutations, 3 missense mutations and one splice region variant (Figure 4.5). Sequencing of matched normal tissues revealed that, while the majority of mutations were sporadic, two APC mutations were germline.

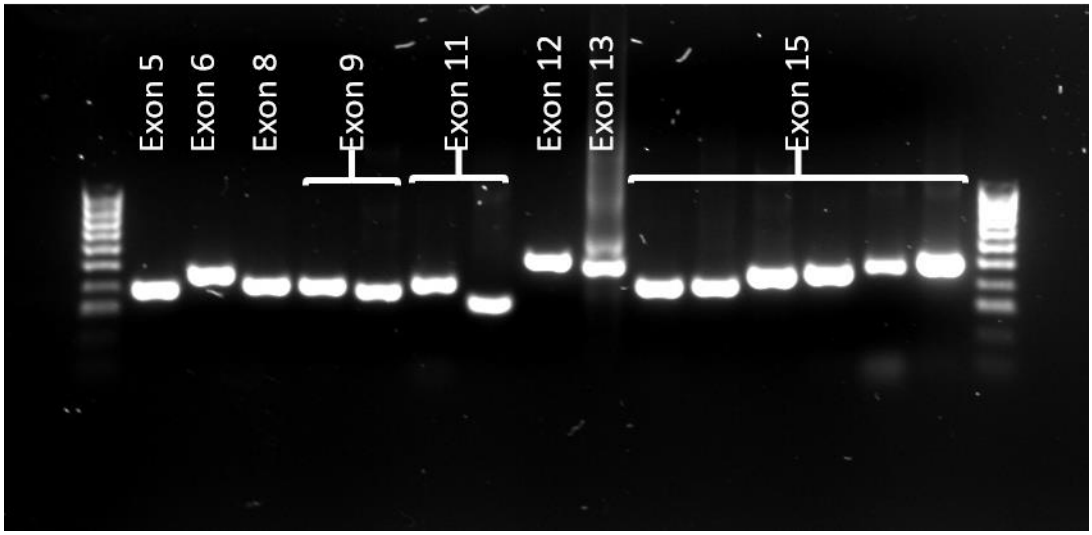

Figure 4.4. PCR amplification of *APC* sequences. Fifteen PCR reactions were used to amplify the regions of *APC* containing the most frequently mutated bases. Two amplicons each were required to span exons 9 and 11 while 6 amplicons spanned exon 15 up to and including the mutation cluster region.

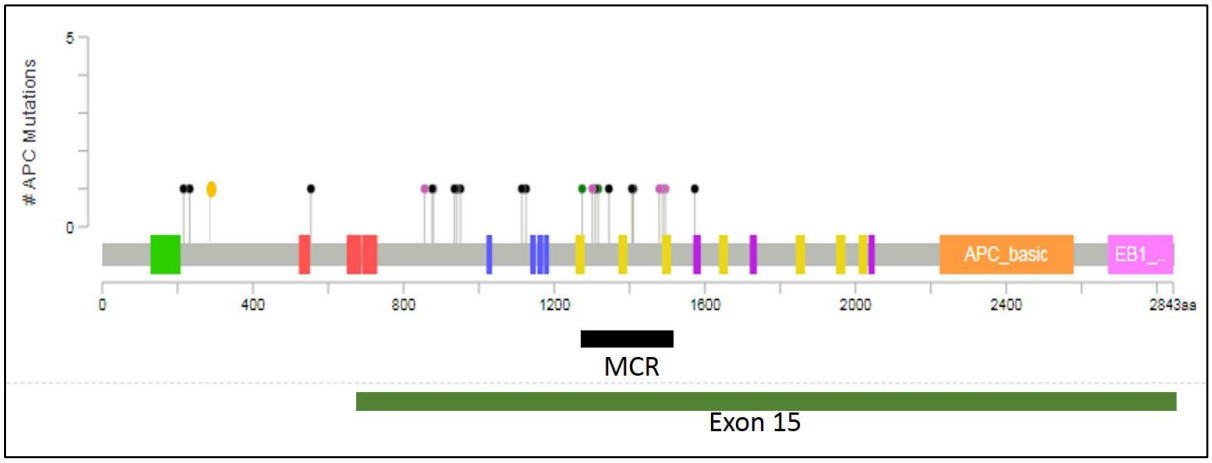

Figure 4.5. *APC* mutations found in EOCRC samples mapped to their position within the protein. Each lollipop represents the position of a mutation with the colours of the lollipop representing different types of mutation: black, nonsense mutations; green, missense mutation; pink, frameshift mutation; yellow, splice variant. While a high proportion of mutations occur within the mutation cluster region (MCR, black bar), mutations are found throughout the first 1600 codons of *APC* including some upstream of exon 15 (green bar). Protein domains: green, oligomerisation; red, armadillo repeats; blue, 15 amino acid repeats; yellow, 20 amino acid repeats; purple, SAMP motif. The plot was generated using Mutation Mapper (www.cbioportal.org) and manually curated.

4.3.5 The two germline *APC* Variants

p.V1125A. A *c.3374T>C* variant, found in one sample in the current cohort, has a low frequency in the general population (ExAC MAF 0.00057). This variant results in a substitution at codon 1125 of a valine to an alanine, and is regarded as a conservative change as they are closely related amino acids. Coupled with the occurrence of the variant in the general population (F. F. Li et al., 2015), as well as its position in a poorly conserved region of the gene, the *c.3374T>C* variant is considered benign.

p.E1317Q. A *c.3949G>C* mutation found in one patient resulted in a glutamic acid to glutamine change at codon 1317 of APC, in a region of the gene poorly conserved between species. The *c.3949G>C* variant is found at low frequency in the general population (ExAC MAF 0.00413; 1000Genomes MAF 0.003). ClinVar has conflicting interpretations of pathogenicity.

In this study this mutation was found in a 36 year old female (Patient 15310) with a stage 2 adenocarcinoma with no lymph node involvement or metastasis. There is no reported family history of bowel cancer. Interestingly, while the matched normal tissue showed the mutation to be heterozygous, the wild type allele was barely visible in the tumour. Moreover, this apparent loss of heterozygosity in the tumour occurred despite the positive staining of the tumour for all four mismatch repair proteins deeming the tumour microsatellite stable. It is reasonable to suggest that this second genetic hit is required for the pathogenicity of the p.E1317Q to occur. This might explain the lack of family history, as the single mutant allele alone may not be sufficient for carcinogenesis, and is an example of the damaging potential caused by of loss of heterozygosity (Figure 4.6).

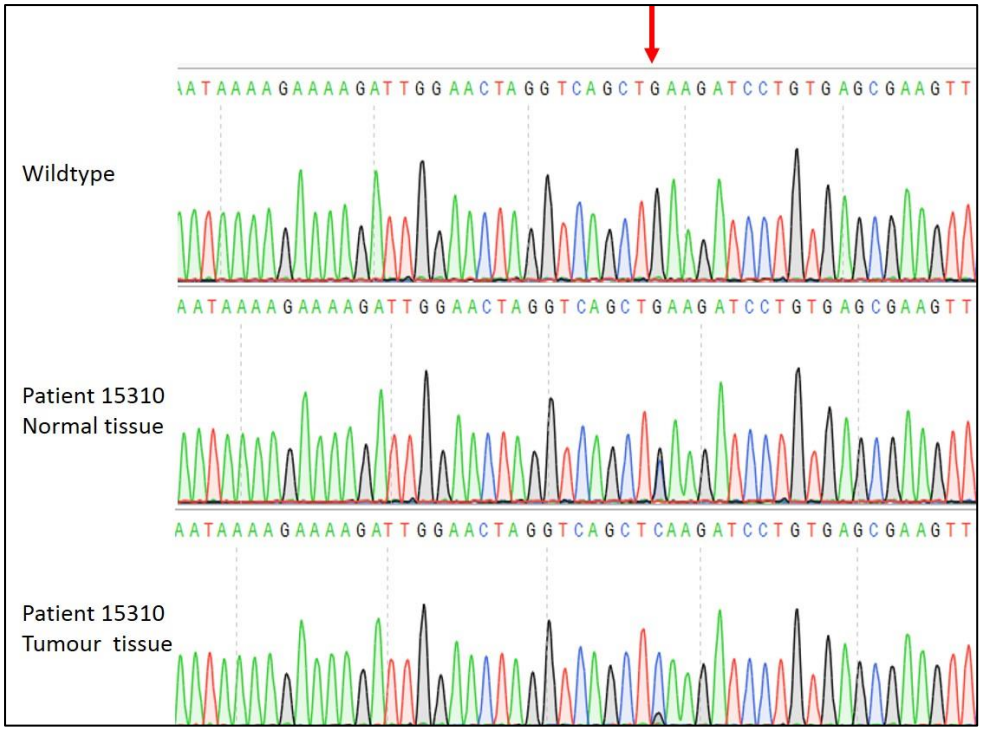

Figure 4.6. Sanger sequencing of *APC c.3949G>C* mutation. The guanine base indicated by the red arrow in the wild type *APC* sequence is mutated to a cytosine in one allele in the normal tissue of the patient to give a heterozygous sequence. In the tumour tissue the wild type guanine allele is barely visible.

4.3.6 Sporadic *APC* mutations

As well as the two germline *APC* mutations, 22 sporadic mutations (20 unique plus one found in two samples) were found in the cohort (Table 4.2) and confirmed by Sanger sequencing. None of these mutations were seen in the matched normal tissue. All but two sporadic mutations were predicted to be damaging, as they caused premature truncation of the gene. The remaining two were missense mutations, both of which were predicted to be benign. Fifteen were previously discovered mutations (references in Table 4.2), with 7 novel mutations identified.

Table 4.2. Sporadic *APC* mutations found in tumour tissue but not in matched normal tissue of EOCRC patients.

Patient Number	Microsatellite Instability	Mutation			Mutational Outcome	Mutational Assessment[1]	Reference
		DNA	rsID	Protein			
10006	MSI	*c.2626C>T*	rs121913333	p.R876X	Nonsense	Damaging	(Powell et al., 1992)
10765	MSS	*c.835-8A>G*	rs1064793022	Splice variant	Truncation[2]	Damaging	(Fostira & Yannoukakos, 2010)
12294	MSS	*c.1660C>T*	rs137854573	p.R554X	Nonsense	Damaging	(Friedl & Aretz, 2005)
12961	MSS	*c.2853T>G*	rs575406600	p.Y951X	Nonsense	Damaging	(Friedl & Aretz, 2005)
13046	MSS	*c.4488delT*	n/a	p.P1497fsX	Frameshift	Damaging	This Study
		c.2805C>A	rs137854575	p.Y935X	Nonsense	Damaging	(Simbolo et al., 2015)
13501	MSS	*c.2636delA*	n/a	p.Q879fsX	Frameshift	Damaging	This Study
13564	MSS	*c.4033G>T*	rs1211642532	P.E1345X	Nonsense	Damaging	(Gomez-Fernandez et al., 2009)
14459	MSS	*c.4463delT*	rs1114167577	p.L1488fsX	Frameshift	Damaging	(Miyaki et al., 1994)
		c.3826delT	n/a	p.S1276fsX	Frameshift	Damaging	This Study
15310	MSS	*c.4216C>T*	rs587782518	p.Q1406X	Nonsense	Damaging	(Lagarde et al., 2010)
16993	MSS	*c.4717G>T*	rs878853217	p.E1573X	Nonsense	Damaging	(Aceto et al., 2005)
		c.2821G>T	n/a	p.E941X	Nonsense	Damaging	(Miyaki et al., 1994)
17871	MSS	*c.4485delT*	n/a	p.S1459fsX	Frameshift	Damaging	This Study
		c.2626C>T	rs121913333	p.R876X	Nonsense	Damaging	(Powell et al., 1992)
18090	MSS	*c.3925G>T*	n/a	p.E1309X	Nonsense	Damaging	(Luchtenborg et al., 2004)
19199	MSS	*c.646C>T*	rs62619935	p.R216X	Nonsense	Damaging	(Gomez-Fernandez et al., 2009)
19513	MSS	*c.3740C>T*	n/a	p.A1247V	Missense	Benign	This Study
		c.3340C>T	rs121913331	p.R1114X	Nonsense	Damaging	(Aceto et al., 2005)
20039	MSS	*c.4230C>A*	n/a.	p.C1410X	Nonsense	Damaging	This Study
20085	MSI	*c.694C>T*	rs397515734	p.R232X	Nonsense	Damaging	(Simbolo et al., 2015)

Table 4.2. (continued) Sporadic *APC* mutations found in tumour tissue but not in matched normal tissue of EOCRC patients.

Patient Number	Microsatellite Instability	Mutation			Mutational Outcome	Mutational Assessment[1]	Reference
		DNA	rsID	Protein			
21025	MSI	*c.2563_2564dupGA*	n/a	p.R856fsX	Frameshift	Damaging	This Study

[1], Missense mutations were considered damaging if 3/5 or more software tools predicted the mutation to be possibly or probably damaging. Nonsense or frameshift mutations were all considered damaging without predictive tools. [2], The splice variant *c.835-8A>G* results in a frameshift that ultimately leads to premature truncation of *APC*. MSS, microsatellite stable; MSI, microsatellite instability; n/a, these mutations have no rsID.

4.3.7 Microsatellite instability

Of the 24 samples for which sections were available, 21 reportedly stained for all four mismatch repair proteins and were therefore considered microsatellite stable. All three of the remaining samples displayed microsatellite instability, with one negative for MSH6 staining and two negative for PMS2 staining. These three tumours did not cluster at any location in the colon, as they were found in the ascending colon, splenic flexure and the sigmoid colon respectively.

Interestingly, one of the 21 samples that stained positively for all four mismatch repair genes was reported to have a *BRAF* mutation. This sample was the only adenosquamous carcinoma in the cohort and also stained weakly for E-cadherin. This combination of a *BRAF* mutation in a microsatellite stable adenosquamous colon cancer has been reported before (Ishida et al., 2017).

4.3.8 Loss of heterozygosity is associated with *APC* mutation near codon 1300

Loss of heterozygosity was determined by the *APC* sequence data, when loss of one allele of a single nucleotide polymorphism was seen in the tumour sequence where two different alleles were seen in matched normal tissue (Section 2.2.3.13). Three out of the 25 samples in the cohort (12%) displayed a loss of heterozygosity. Of particular note was that these three samples all had *APC* mutations clustered around codon 1300, a phenomenon previously reported in sporadic CRC (Rowan et al., 2000).

4.4 Discussion

In contrast to the cohort of early-onset CRCs from Pakistan (Chapter 3) that were primarily characterised by the presence of signet rings, the local cohort predominantly carried colorectal adenocarcinomas. We found one tumour in our cohort of EOCRC tumours with evidence of a previously described *CDH1* germline mutation. That, combined with evidence of E-cadherin staining in all but the single SRCC case within the cohort, leads to the conclusion that loss of E-cadherin is not a major factor in the pathogenesis of early-onset CRCs in New Zealand. This finding is in agreement with a recent study where commercially available 46- or 50-gene panels for Next Generation Sequencing also failed to find *CDH1* to be commonly mutated in early-onset CRCs, or indeed at any age of onset (Willauer et al., 2019).

The one tumour with evidence of a *CDH1* germline mutation was found in a 34 year old male. This was a *c.88C>A* variant reportedly present at low frequency in the general population (https://www.ncbi.nlm.nih.gov/clinvar/variation/127933/). However, the very young age of diagnosis and the presence of mucin in the tumour suggests a similarity to hereditary diffuse gastric cancer patients, where *CDH1* mutations are causative (Guilford et al., 1998). That, along with the co-occurrence of *c.88C>A* in other diseases such as lobular breast and gastric cancer (Masciari et al., 2007; Molinaro et al., 2014) and the observation that c.88C>A affected the localisation of E-cadherin *in vitro* (Vogelaar et al., 2013), suggests that *c.88C>A* may be causative in this individual. Although the presence of strong E-cadherin staining at the cell membranes in the tumour reduces this possibility, the presence of this mutation may indicate familial risk and screening of family members should be considered.

Only one sample was histologically classified as a SRCC and interestingly this was the same tumour that was classified negative for E-cadherin staining, providing further evidence that loss of this inter-cellular protein is an important step in the development of SRCC tumours. One other tumour also showed partial loss of E-cadherin. The absence of any *CDH1* mutation in these two samples with reduced or absent E-cadherin expression suggests other forms of silencing of this gene are likely to have occurred. Methylation-induced silencing of *CDH1* has been found in colorectal cancers (Lee et al., 2004; Wheeler et al., 2001), including SRCCs where loss of E-cadherin was associated with an increase in expression of MeCP2, a methyl-binding protein (Darwanto et al., 2003). Promoter methylation of one allele as a second genetic hit may occur alongside genetic loss of the other allele as has been described for *CDH1* in gastric cancer (Oliveira et al., 2009), providing additional argument for *CDH1* methylation as a possible mechanism in the current cohort. Time constraints precluded methylation analyses of our cohort but this is considered a priority for future research.

In contrast to the observed lack of mutations in the *CDH1* gene, mutations of *APC* were identified in 18/25 (72%) of early-onset CRC cases in our cohort. This is higher than some other studies looking at *APC* mutation in early-onset CRCs (Kothari et al., 2016; Willauer et al., 2019) that may, in part, reflect the methodology used here. The *APC* gene is long, encompassing 8529 bases of coding sequence over multiple exons covering over 100 kilobases of chromosome 5, making sequencing strategies cumbersome and expensive. Therefore, many studies have focused on mutation hotspots in the gene, particularly the mutation cluster region between amino acids 1286 and 1513 (Miyoshi et al., 1992) or more widely within the final,

large exon which encompasses the majority of *APC* mutations. This has the potential to miss mutations that fall outside these regions.

We developed a sequencing strategy to maximize the mutations identified while maintaining a manageable, cost-effective amount of sequencing. Firstly, regions of *APC* with very low rates of mutation in sporadic CRCs were identified using the Catalogue Of Somatic Mutation In Cancer (COSMIC) database, and excluded from our sequencing strategy. That left exons 5, 6, 8, 9, 11, 12 and 13, as well as exon 15 from amino acids 788 to 1593. These exons encompass the vast majority of non-synonymous mutations of the APC gene in the COSMIC database, and each was sequenced. This approach identified 24 mutations in 18/25 (72%) tumour samples across these regions of the *APC* gene. Importantly, if we had only sequenced the mutation cluster region, this number would have dropped to 8/25 (32%). Likewise, mutations in the final exon of the gene were identified in only 14/25 (56%) of the samples. These findings convincingly highlight that protocols applied to study *APC* mutations are of critical importance, and that studies limited to "hotspot" regions of the gene may significantly under-report the mutation rate.

A recent study investigating mutations in two separate cohorts of colorectal cancer gave considerably different rates of *APC* mutation between the cohorts (Willauer et al., 2019) that were investigated using different screening panels. For patients under 50 years of age, the *APC* mutation rates of the tumours were 41.6% and 65.2% for the two cohorts. The commercial panel that identified the lower rate of APC mutations only investigated mutations within the last exon of *APC*, mostly within the mutation cluster region. The second cohort was studied using a mixture of platforms, suggesting the method used to identify *APC* mutations and the amount of the gene covered has a large impact on the rate of mutations reported. In comparison, the rate of tumours with mutations in the final exon of APC within our cohort was 56% (14/25). Of further interest was that, with the exception of patients under 30 years old, the rates of *APC* mutation in the two cohorts detailed in the Willauer study were not very different between early-onset and late-onset CRCs (Willauer et al., 2019). Only one patient in our cohort was aged under 30, a male aged 28 with left sided adenocarcinoma. Of note this patient did not have an *APC* mutation.

Of interest was our finding of two germline mutations in our cohort, despite familial adenomatous polyposis being one of the exclusion criteria. Perhaps unsurprisingly, while 21/24 mutations were either nonsense or frameshift mutations resulting in truncated proteins, both

the germline variants were missense mutations. The p.V1125A variant appears to be benign, whereas the p.E1317Q variant that replaces a glutamic acid residue with a glutamine has previously been found to be significantly associated with cancer with an odds ratio of 11.17 in a study of 164 unrelated colorectal adenoma patients (Lamlum et al., 2000). While glutamic acid and glutamine are structurally similar, this alteration may have the potential to cause a change in the charge at that specific amino acid position that may affect folding or some other aspect of protein regulation or function.

The finding that the wild type allele, while present in the patient's normal tissue, was almost completely absent in the tumour of the patient with the germline p.E1317Q mutation, suggests that p.E1317Q acts as a susceptibility allele, only becoming pathogenic in the absence of its wild type partner.

This is further evidenced in a study of a cancer-prone family in Scotland, where two family members with colon cancer carried the p.E1317Q mutation. In tumours from both those individuals, the wild type allele was lost (White et al., 1996). Interestingly, the patient in our cohort with evidence of this mutation presented with adenocarcinoma with adenoma, but no family history of adenomas, seemingly ruling out Familial Adenomatous Polyposis. However, a study that detailed the family histories of four unrelated CRC patients exhibiting the same p.E1317Q mutation, found a spectrum of risk that ranged from no known history of colorectal cancer or related disease in one patient, to two of the patients having a first-degree relative with rectal cancer (Frayling et al., 1998). This is further compounded by evidence that the number and rate at which polyps developed in these four patients was widely divergent, where two had 17 and 5 metaplastic polyps respectively, while the other two had none but developed 1 and 3 polyps in 14 and 15 years of follow-up, respectively (Frayling et al., 1998).

These conflicting presentations suggest that p.E1317Q has wide phenotypic variation and may or may not indicate familial risk for relatives carrying the mutation. In light of this, screening of families identified with the p.E1317Q mutation may identify polyp formation or cancer at an earlier stage than otherwise likely.

Furthermore, the finding of a total of three samples with loss of heterozygosity, by the loss of one *APC* allele in the tumour tissue, including the patient described above, was notable in that all three samples had APC mutations clustered near codon 1300, namely codons 1309, 1317 and 1345. While this phenomenon was first identified in familial adenomatous polyposis patients (Lamlum et al., 1999) it was subsequently identified in sporadic CRC (Rowan et al.,

2000). Similarly, the three samples we found with LOH associated with mutations near codon 1300 include one germline and two sporadic mutations. Although we did not investigate the mechanism of LOH in these samples, whether due to specific loss of an APC allele or a wider loss of a chromosomal region or arm, our findings of LOH in samples with mutations around codon 1300 add further evidence to this intriguing observation and expand it to include early-onset CRC.

In a review of 1025 CRC patients under 50 years of age, tumours were found to be predominantly distal (49.1% rectal, 29.1% left-sided) with a minority (21.9%) being right sided (Dozois et al., 2008). Similarly, the current cohort was generally distal (68%) with almost half (48%) being located in the sigmoid colon. There were some gender differences in location with 6/11 (54.5%) of males having distal tumours in contrast to 11/14 (78.6%) of females. Despite our cohort ranging in age from 28 to 50 there was no tendency for younger patients to have distal cancers with the mean age of those with proximal cancers being 40.7 years as opposed to 44.5 years for those with distal tumours. Interestingly, although 17/25 (68%) of the cohort had distal tumours, among those tumours with *APC* mutations, 11/18 (61.1%) were distal while 6/7 (85.7%) of tumours lacking *APC* mutation were distal.

Levels of microsatellite instability have been reported to be higher in early-onset CRC than in older patients with rates amongst patients aged under 50 years of age ranging between 19.8% and 31% (Cheong et al., 2019; Perea et al., 2010) while older patients display lower rates (Cheong et al., 2019). In our cohort we found only 3/25 (12%) of cases with microsatellite instability. The lower rate may be partly attributed to the exclusion of Lynch syndrome patients, in whom a germline mutation in a DNA mismatch repair gene results in a defective mismatch repair process. This is exemplified by an unrelated study of microsatellite instability in CRC patients aged under 45 years old, where 11 (24%) of the 45 patients had microsatellite instability but when patients with germline mutations in mismatch repair genes were removed, this reduced to 6/40 (15%) (Salovaara et al., 2000). We saw no association of microsatellite instability with tumour location, with the three microsatellite instability tumours being found in the ascending colon, splenic flexure and sigmoid colon respectively.

When EOCRC patients are further stratified by age, it is clear that the younger the age at diagnosis, the higher the likelihood of microsatellite instability. Studies looking at younger age groups show higher rates of microsatellite instability that include 40.5% in a cohort of under-40s (J. T. Liang et al., 2003), 47.5% in under-30s (Farrington et al., 1998) and 72.7% in under-

25s (Durno et al., 2005). The median age of the cohort in our study is 44 years which may additionally explain the lower microsatellite instability rate. In support of this hypothesis is the finding that two of the six patients (33%) under the age of 40 in the cohort displayed microsatellite instability.

Like *CDH1*, *APC* promoter methylation has long been associated with colorectal cancer (Esteller et al., 2000). With the high rate of *APC* mutation in our cohort, it is possible that promoter methylation may be involved as the second-hit in some of these cases. Unlike *CDH1*, where most mutations are loss-of-function mutations, the truncated APC proteins that arise from many of the mutations still retain some functionality (Chandra et al., 2012; Christie et al., 2013), and indeed the altered functionality of the truncated forms appears to promote carcinogenesis (Albuquerque et al., 2002). Promoter methylation, which results in loss of expression, would preclude a silenced allele from performing these altered functions, thus it may be that methylation of the *APC* gene occurs in cells where the other allele is already truncated and performing the oncogenic role required by the cells. Allelic loss of *APC* has been shown to occur more often in tandem with a truncation at around codon 1300 (Rowan et al., 2000) and accordingly it may be that protein loss via promoter methylation may occur alongside similarly positioned *APC* truncations.

4.4.1 Conclusions

The absence of any pathogenic *CDH1* mutations in all the adenocarcinoma samples in our early-onset CRC cohort suggests that, unlike the early-onset SRCC tumours from Pakistan, mutation of this gene is not a common cause of CRC in under 50 year-olds in New Zealand. This is further evidenced by the strong E-cadherin staining in all the adenocarcinoma samples tested. This includes strong E-cadherin staining in the sample with a p.P30T variant, which is in keeping with the bulk of evidence in the literature that p.P30T is not pathogenic. The negative E-cadherin staining in the sole SRCC sample concurs with our conclusions from the previous chapter that loss of E-cadherin is, however, a common factor in the development of colorectal SRCC, even when due to other mechanisms than *CDH1* mutation. The weak E-cadherin staining found in the only adenosquamous tumour in our cohort may indicate a different aetiology to the more common adenocarcinomas.

The presence of a high number of *APC* mutations in our cohort compared to other studies of early-onset CRC may reflect the methodology used, particularly the extent of the gene that is

investigated in different studies and that, particularly with recent advances in sequencing technology, greater coverage of the *APC* gene is desirable when looking for possibly pathogenic *APC* mutations. Whereas the association of loss of heterozygosity at the *APC* locus with mutations around codon 1300 is not a novel discovery, it does add to existing evidence around this phenomenon and, importantly, expands it to include early-onset CRC.

Chapter 5 Gene editing of mutant *APC* in HT29 cells

5.1 Introduction

While Chapter 3 described a role for E-cadherin silencing in the development of early-onset colorectal signet-ring cell carcinomas, Chapter 4 showed that this was not as prevalent in the cohort of early-onset, colorectal adenocarcinomas, where only the sole SRCC case showed negative staining for E-cadherin. Instead, the finding of high rates of *APC* mutation in this cohort suggests that early-onset colorectal adenocarcinomas in our community may have similar aetiologies to tumours in older patients, where there is considerable existing evidence of *APC* mutation as an early event in carcinogenesis (Powell et al., 1992).

While current thinking is that the development of sporadic cancers may be driven by environmental and/or lifestyle factors (J. Keenan et al., 2017; Richardson et al., 2016), there is also increasing evidence to suggest a role for an individual's gut microbiota in the aetiology of these cancers. Specifically, long-term carriage of toxin-producing strains of gut bacteria is considered to have the potential to initiate colorectal carcinogenesis (Tjalsma et al., 2012).

Bacteroides species are highly abundant in the gut microbiota (Wick & Sears, 2010) and are generally considered to contribute to human health (Mazmanian et al., 2005). However, select strains of *B. fragilis* can express a toxin capable of triggering colonic inflammation in mice (Rhee et al., 2009) and damaging cellular DNA (Goodwin et al., 2011). Carriage of these enterotoxigenic strains of *B. fragilis* (ETBF) is reportedly increased in patients with colorectal cancer (J. I. Keenan et al., 2016; Toprak et al., 2006), and long-term colonic carriage of ETBF is associated with significant risk of low-grade colonic dysplasia (Purcell et al., 2017).

To date, *in vitro* studies of mechanisms that might underlie *B. fragilis* toxin (BFT)-mediated carcinogenesis have been performed using the human colonic cancer HT29 cell line or its derivative HT29/c1 (Huet et al., 1987). Numerous studies have demonstrated the ability of purified BFT to cause these cells to round up, reportedly as a result of toxin-mediated cleavage of extracellular E-cadherin (Sears, 2009). However, when BFT is added to the human colorectal carcinoma HCT116 cell line, these cells retain their normal morphology.

The *APC* gene in HT29 cells carries mutations that allow expression of only truncated APC, whereas the gene in HCT116 cells expresses full length wild type APC (El-Bahrawy et al., 2004) (Figure 5.1), raising the possibility that the reported "exquisite sensitivity" of HT29 cells

to BFT may reflect their *APC* genotype rather than toxin-mediated loss of E-cadherin *per se*. If so, using cell culture models where the cells are already partly along the pathway to cancer (and drawing conclusions regarding the effects of environmental factors on colonic epithelial cells) has the potential to misrepresent the risk long-term carriage of ETBF has on the development of carcinogenesis in an infected individual. As such, HT29 cells may not represent a useful cell line in which to model the effects of BFT on normal epithelial cells in the gut, in order to better understand how this bacterial toxin might drive colon carcinogenesis.

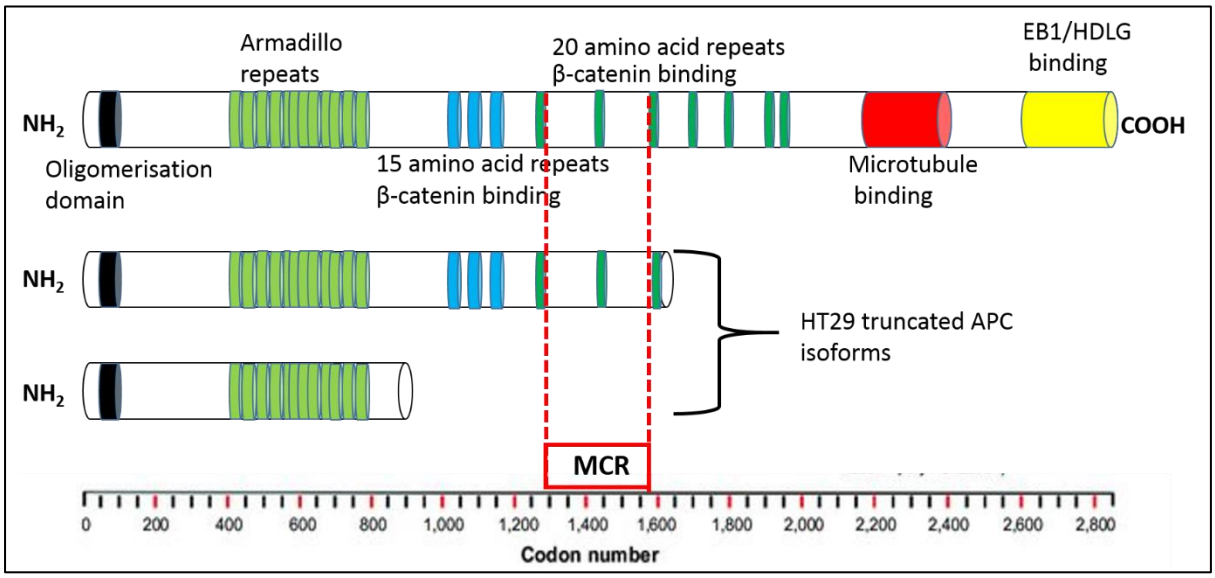

Figure 5.1 APC functional domains and isoforms in HT29 cells. Top: Full length wild type APC protein with functional domains annotated. Middle and bottom: Truncated APC protein isoforms found in HT29 cells. The shortest truncated APC at the bottom is the isoform corrected in this chapter. MCR, mutation cluster region.

One option is to study HCT116 cells as they carry an intact *APC* gene. HCT116 cells, however, also carry an activating mutation in the *CTNNB1* gene, which encodes β-catenin and, like cells with mutated *APC*, also have a defective Wnt signaling pathway (El-Bahrawy et al., 2004).

An alternative option is to express full-length APC protein in HT29 cells and use these cells to measure the effects of the toxin on colonic epithelial cells with correctly regulated Wnt signalling pathways. APC has been expressed in HT29 cells previously (Morin et al., 1996), with the inducible APC-expressing cells used to elucidate downstream targets of Wnt signaling (T. C. He et al., 1998). However, Morin and colleagues used a zinc-inducible promoter to

express APC. The *B. fragilis* toxin is a zinc metalloprotease (Moncrief et al., 1995), meaning the use of a zinc-inducible promoter was not an option for our study. Instead we sought to edit the *APC* gene in HT29 cells using a CRISPR-Cas9n (nickase) approach (Ran et al., 2013) as a means to enable wild type APC expression from the HT29 cells' genomic DNA.

While both *APC* alleles are mutated in HT29 cells it was felt that maintaining one mutant allele would be the prudent option in the first instance as truncated APC isoforms are known to be critical for the proliferation of CRC cells (Chandra et al., 2012). In other words the removal of both truncated forms may have been too detrimental to the cells. Editing of the second allele could be considered after the establishment of a stable cell line with one corrected *APC* allele. It was decided to correct the c.2557G>T mutation which would leave the c.4666_4667insA mutation unaltered, thereby leaving a truncated isoform with β-catenin binding domains that may be critical in the function of the truncated isoform (Figure 5.1).

This chapter details the development of a derivative of the HT29 cell line expressing full-length APC as a model for investigating a role for BFT in the initiation of colon carcinogenesis.

The aims of the chapter were to:

1. Edit genomic *APC* using a CRISPR-Cas9 system to allow endogenous expression of wild type APC from HT29 cells.
2. Characterise the resultant cell line in comparison with the original HT29 cell line.
3. Compare the effects of BFT on HT29 cells with and without full-length APC expression.

5.2 Methods

CRISPR-Cas9n editing of the mutated *APC* gene in HT29 cells was carried out using the methods described in Chapter 2.2.4. Briefly, plasmids carrying antibiotic resistance genes as well as containing the gene for the Cas9 nickase enzyme, targeted to the *c.2557G>T* mutation in the *APC* gene, and a template fragment encompassing the relevant genomic region of wild type *APC* to copy, were transformed into HT29 cells and cultured in the presence of selective antibiotics. Colonies of surviving cells were then tested to identify clones with *APC* successfully edited at the c.2557 locus. Putative edited clones were confirmed by both Sanger and next-generation sequencing. The other mutant allele, *c.4666_4667insA*, was not corrected.

The growth rate of a clone containing corrected *APC* was compared to those of the parent HT29 cell line and the colonic epithelial cell line HCT116, both of which have defective Wnt signalling, caused by mutated *APC* (HT29) or β-catenin (HCT116). This was done by growing the cell lines in appropriate media for 96 hours at 37°C, and counting the number of cells per well at 24 hour intervals (Section 2.2.2.2).

Immunofluorescence microscopy was used to show the expression and localization of E-cadherin, β-catenin and actin, while protein expression was investigated by Western blotting of cell lysates, using materials (Table 2.7) and methods (Section 2.2.4.12) as described in Chapter 2.

Immunoprecipitation of the APC protein in cell lysates was carried out using specific N- and C-terminal antibodies (section 2.2.4.11). Precipitated proteins were run on 4-20% gradient polyacrylamide gels and probed with antibodies raised against the C- and N-terminals of APC respectively.

5.3 Results

5.3.1 HT29 cells contain three alleles encoding truncated forms of the APC protein.

A previous study utilizing spectral karyotyping to assess chromosomal abnormalities amongst commonly studied colonic epithelial cells found that there were three copies of the long arm of chromosome 5 in the HT29 cell line (Abdel-Rahman et al., 2001). HT29 cells are known to exhibit two truncated forms of *APC*, a single base pair change, *c.2557G>T*, introducing a stop codon resulting in a p.853X truncation, and a single base pair insertion, *c.4666_4667insA*, resulting in a p.T1556fsX3 frameshift and subsequent truncation.

To date there are no reports of relative allelic ratios of these truncations in light of the additional chromosome. In the previous chapter, next generation sequencing was described as a way to screen the APC gene for evidence of mutations in patient samples (Section 2.2.3). Using a similar approach to sequence the APC gene in the HT29 cells allowed an assessment of the relative proportions of each mutant allele within HT29 cells. Accordingly, when HT29 cells were sequenced in this way, 1477/2166 (68%) of reads for the *c.2557G>T* mutation exhibited the *G* allele while the remaining 689/2166 (32%) displayed the mutant *T* allele. We concluded

therefore that 1/3 alleles contain the *c.2557G>T* mutated APC, while 2/3 alleles contain the *c.4666_4667insA* mutation.

5.3.2 CRISPR-Cas9n editing can incorporate errors into target sequences.

A CRISPR-Cas9n approach was used to correct the c.2557G>T mutation in the APC gene of HT29 cells. The CRISPR-Cas9n editing process can however result in both on (or near)-target and off-target errors. Accordingly, Next Generation Sequencing was used again to confirm that CRISPR-Cas9n editing of selected clones resulted in clones with just the c.2557G>T mutation corrected. The importance of including this step is highlighted by the finding that one of the selected clones had a 24 base pair deletion proximal to the mutated base in *APC* (Figure 5.2).

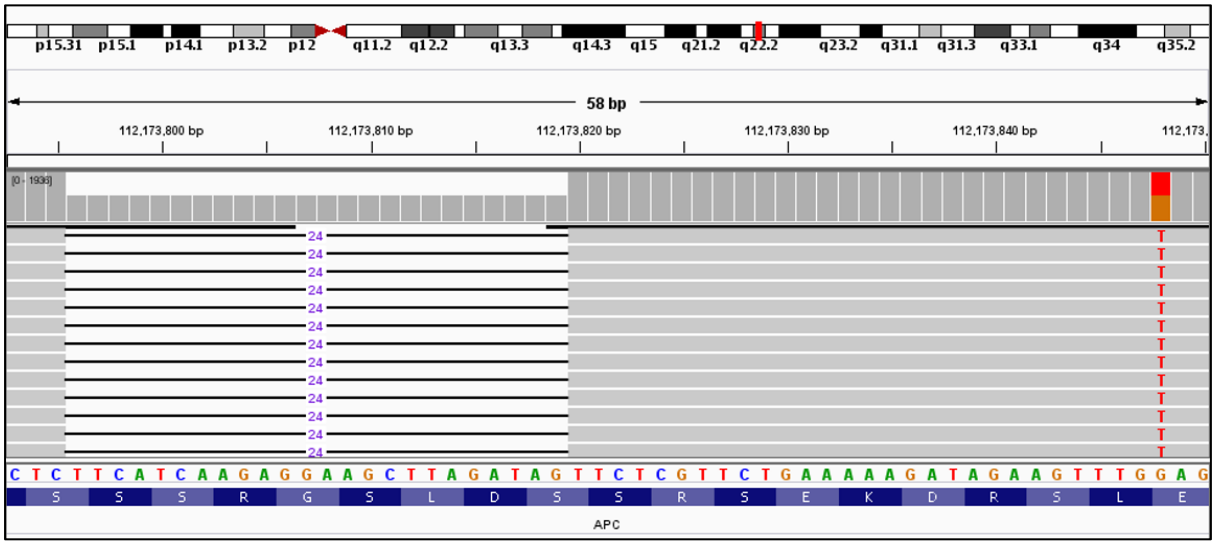

Figure 5.2. Short deletion incorporated during CRISPR editing of *APC* gene. Sequence reads as viewed on the Integrative Genomics Viewer showing the failure of CRISPR-Cas9n to incorporate the correct sequence in one of the 96 selected clones. The top of the figure shows chromosome 5 with the highlighted region of 5q22.2 containing the *APC* gene. Below that the mutant thymine, highlighted on the right of the sequence reads, has not been edited while a 24 base pair sequence has erroneously been deleted upstream of the mutation.

5.3.3 Confirmation of *APC* editing in HT29

Once individual clones from the CRISPR-Cas9n editing experiment (described above in Section 5.2) had grown to approximately 80% confluency in wells of a 96-well plate, the cells were harvested. For each clone some cells were seeded into a new plate, while others were used to confirm that editing of the *APC* gene had occurred.

Cells from 96 individual CRISPR-Cas9n edited clones were centrifuged, washed and resuspended in 10mM Tris-CL pH8, boiled for 15 min and centrifuged. The supernatant was then used for PCR of the *APC* sequence spanning the mutated base at *c.2557*, with XhoI digestion of the PCR product (see Section 2.2.4.5) to confirm the incorporation of template DNA into the HT29 genome (Figure 5.3). Sanger sequencing of one clone, referred to as Clone 30, was carried out to confirm incorporation of the corrected base and restriction sites (Figure 5.4). This was later confirmed by Next Generation Sequencing as described in Section 2.2.3 (not shown). Due to its confirmed genomic correction, Clone 30 was subsequently referred to as HT29 $APC^{+/-}$.

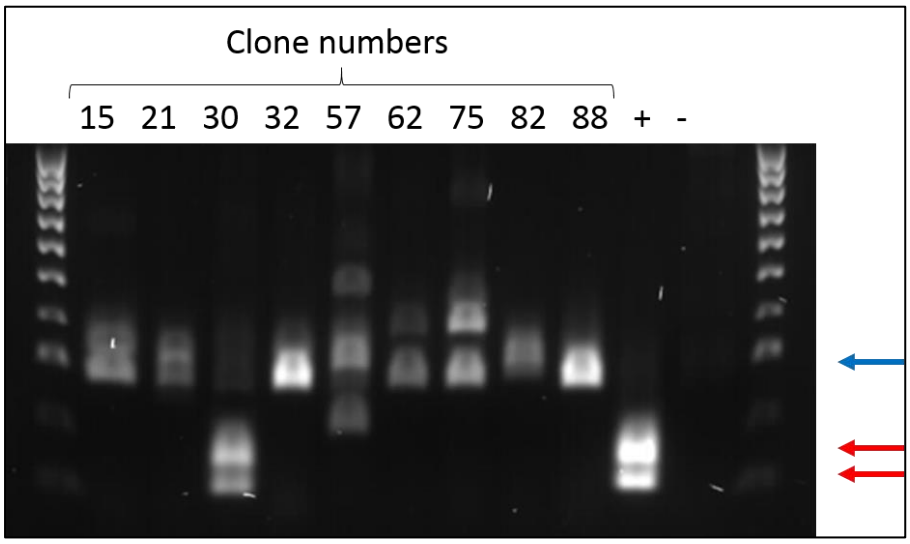

Figure 5.3. Digestion of PCR products from CRISPR-Cas9n edited clones. PCR products spanning the edited region of the *APC* gene were digested with XhoI. The undigested PCR product is indicated by the blue arrow, with the two XhoI digestion products obtained from Clone 30 indicated by red arrows. Template plasmid DNA was used as a positive control (+) and water was used as a no template control (-) in the PCR reactions.

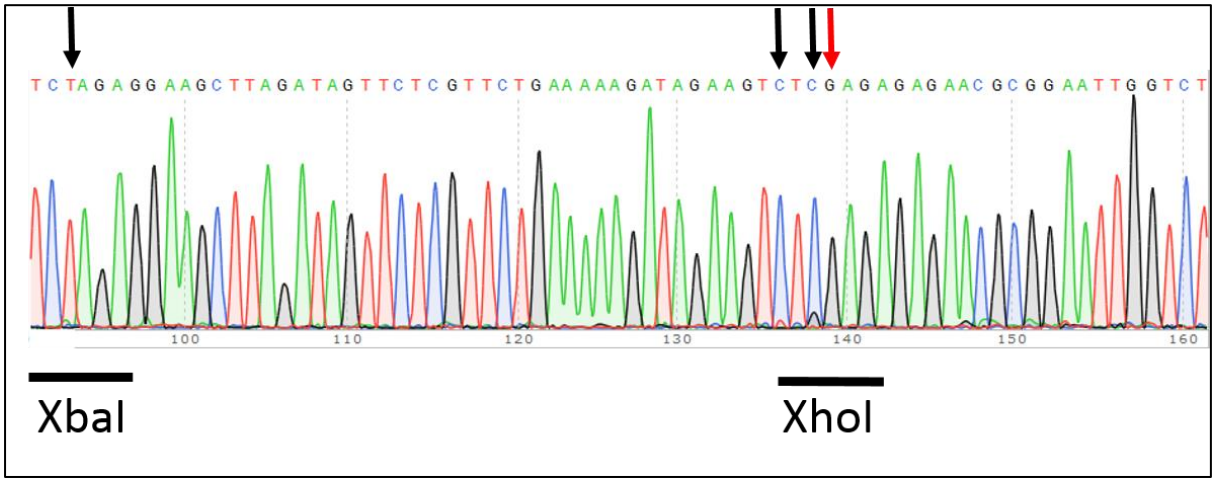

Figure 5.4. Sanger sequence of *APC* confirming incorporation of template fragment correction of the *c.2557G>T* mutation. The *APC* sequence of Clone 30 shows incorporation of XbaI and XhoI sites indicating that the Cas9n enzyme has incorporated the template fragment DNA into the genomic *APC* DNA. Black arrows indicate bases altered to incorporate restriction sites; red arrow indicates the base edited from the mutant thymine to the wild type guanine base.

5.3.4 Correction of the *APC* gene in HT29 cells slows their growth rate

HT29 cells and HT29 $APC^{+/-}$ cells (derived from Clone 30 and shown to have one corrected *APC* allele) (5 x 10^4 cells) were grown in 1 mL of media in 24 well plates for four days. Their growth rate was assessed at 24 hour intervals by lifting and counting the number of cells per well. HT29 cells showed rapid growth as expected for poorly differentiated cancerous cells. However, the HT29 $APC^{+/-}$ cells showed a growth rate that was significantly slower than that of their parent cell line (Figure 5.5).

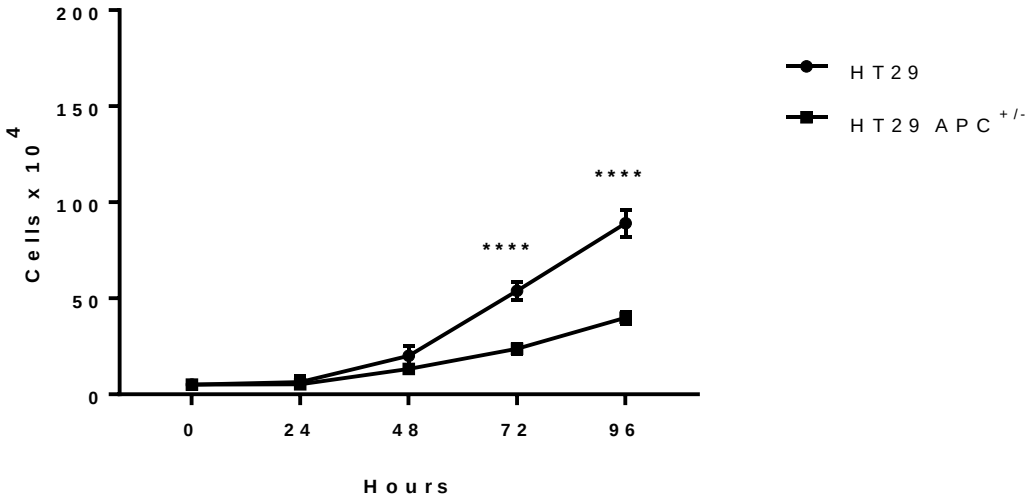

Figure 5.5. Growth rates of HT29 and HT29 *APC*$^{+/-}$. Cell counts were measured over 96 hours at 24 hour intervals for each cell line with the CRISPR-edited cell line showing a lower growth rate than its parent cell line HT29. The results shown are from three independent experiments, analysed using Two-Way ANOVA with Sidak's multiple comparisons test. ****; P<0.0001

Interestingly, the number of non-viable cells, as determined by trypan blue staining (Section 2.2.2.2), was significantly higher in HT29 *APC*$^{+/-}$ than in the parent HT29 cell line (Figure 5.6) suggesting that increased cell death may underlie the observed slower growth rate.

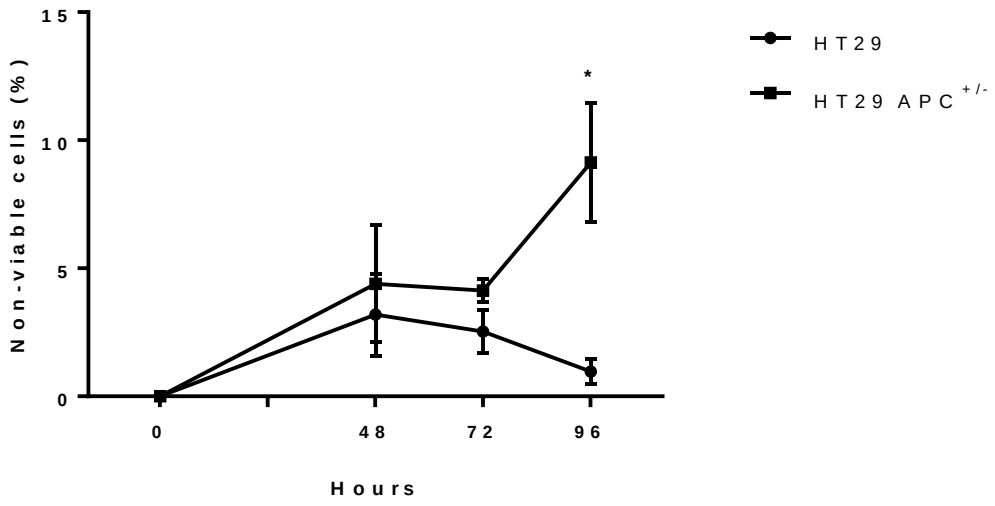

Figure 5.6. Percentage of non-viable cells during growth of cell lines. Dead cells stained with trypan blue were counted and were calculated as a proportion of total cells. A greater proportion of cells with corrected *APC* stain with trypan blue than HT29 cells. The results shown are from three independent experiments, analysed using Two-Way ANOVA with Sidak's multiple comparisons test. *; $P<0.05$.

5.3.5 HT29 *APC*$^{+/-}$ cells maintain morphology after incubation with BFT-enriched supernatant.

Based on the growth curves in Figure 5.5, cells of each of the three cell lines (HT29, HT29 *APC*$^{+/-}$ and HCT116) were seeded in 24 well plates at concentrations between 5 x10^4 and 1 x 10^5 cells/well with the aim of reaching similar levels of confluence after 48 hours of growth. Concentrated enterotoxigenic *B. fragilis* broth culture supernatants, prepared by ammonium sulphate precipitation of broth following bacterial culture (Section 2.2.5.1) and demonstrated to have measureable toxigenic activity (Section 2.2.5.2) were then added to the cells. After 6 hours of incubation with the BFT-enriched supernatant the HT29 cells showed distinctive rounding in response to the toxin, as reported elsewhere (Sears, 2009), while the concentrated NTBF supernatant (used as a negative control) had no effect on cell morphology over the same period of time (Figure 5.7).

Likewise little to no rounding occurred in HT29 APC$^{+/-}$ cells in response to the concentrated ETBF supernatant, with the morphology of the toxin-treated cells observed to closely resemble that seen in HCT116 cells (Figure 5.7).

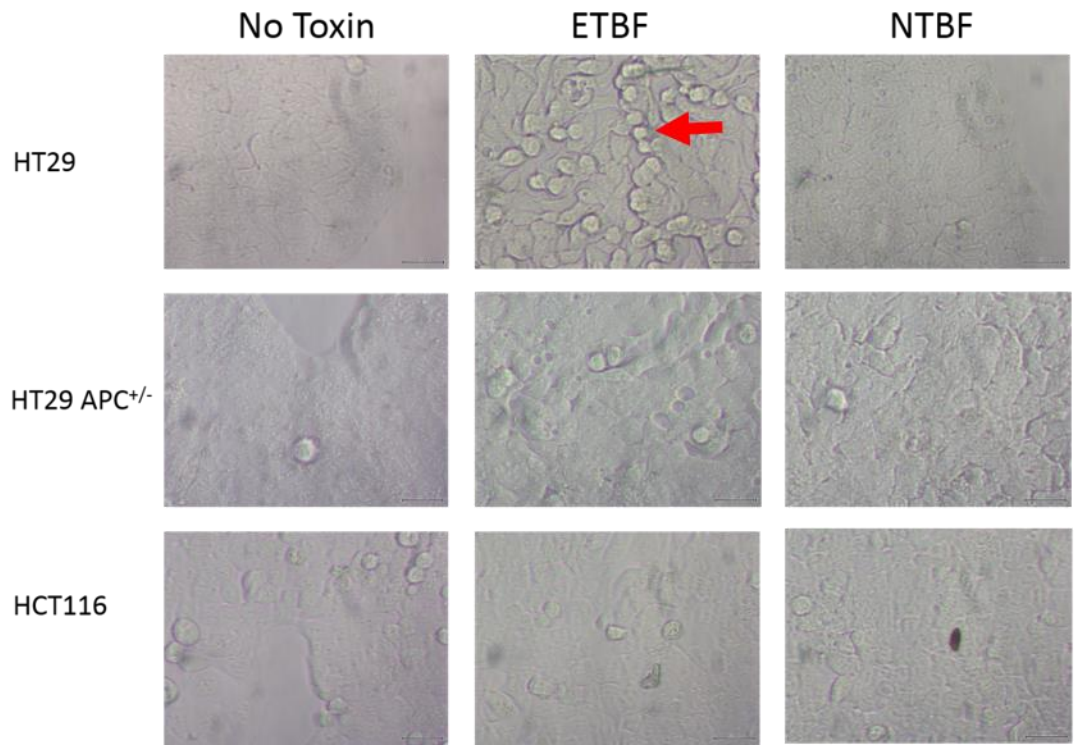

Figure 5.7. Cell morphology following incubation with BFT-enriched ETBF supernatant. The effect of BFT on HT29 cells can be seen by the rounding up of cells (arrowed), whereas no rounding up occurs with supernatant from NTBF. Cellular morphology is barely affected in both HT29 APC$^{+/-}$ and HCT116 cells. Images are representative of three independent experiments. BFT, *B. fragilis* toxin, ETBF, enterotoxigenic *Bacteroides fragilis*; NTBF, non-toxigenic *Bacteroides fragilis*. The cells were viewed under 40x magnification.

5.3.6 *Bacteroides fragilis* toxin-mediated loss of E-cadherin occurs in HT29 *APC*$^{+/-}$ cells

HT29 and HT29 *APC*$^{+/-}$ cells were grown on coverslips for 48 hours (Section 2.2.2.5), after which ETBF or NTBF supernatant was added to the medium and the cells were incubated for a further 6 hours. The cells on coverslips were, fixed, permeabilised, washed and incubated with an anti-E-cadherin antibody overnight at 4°C, before incubation with an AlexaFluor-tagged secondary antibody directed to the E-cadherin antibody (Section 2.2.2.5) for 2 hours at room temperature. Cells were also incubated overnight in Texas-Red Phalloidin, a fluorescently-tagged toxin that binds to actin. Nuclei were then stained with Hoechst 33342 for 30 minutes, and coverslips were washed and mounted on slides for fluorescence microscopy.

E-cadherin was cleaved in both HT29 and HT29 *APC*$^{+/-}$ cells after incubation with the toxin-enriched ETBF supernatant but the loss of actin staining was much greater in HT29 cells than in HT29 APC$^{+/-}$ cells (Figure 5.8). This finding supports the idea that full-length APC has a role in maintaining the integrity of the actin cytoskeleton following exposure to BFT, via its association with the cytoskeleton (Moseley et al., 2007; Narayan & Roy, 2003) and thus cell morphology, as seen in Figure 5.7.

Collectively these results support the hypothesis that toxin-mediated loss of epithelial cell morphology is largely due to the absence of full-length APC protein as opposed to simply loss of inter-cellular E-cadherin.

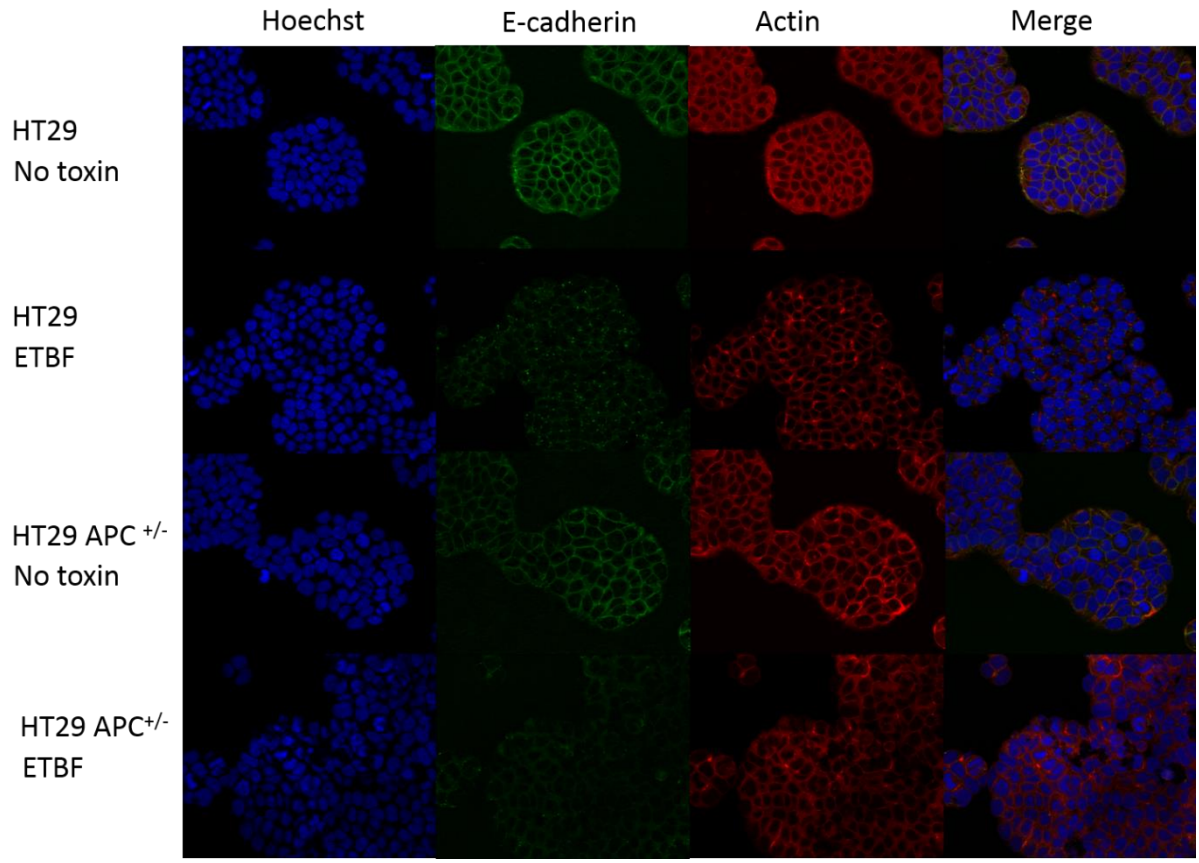

Figure 5.8. Immunofluorescence staining of HT29, HT29 APC$^{+/-}$ cells with and without BFT-enriched supernatant treatment. E-cadherin staining (green) is markedly reduced in both cell lines following incubation with BFT-enriched supernatant. Actin staining (red) reduces in HT29 cells with BFT-enriched supernatant but is maintained in HT29 APC$^{+/-}$ cells. Nuclei are stained blue. The cells were viewed under 40x magnification and the images are representative of three separate experiments.

5.3.7 Full length APC is not identifiable in Western blots of HT29 *APC*$^{+/-}$.

HT29, HT29 *APC*$^{+/-}$ and HCT116 cells were grown to confluency and Western blots were performed on cell lysates to identify the APC protein. Probing with an antibody raised against the N-terminal region of APC produced a band at the expected level in the HCT116 cell lysate but not in either the HT29 or HT29 *APC*$^{+/-}$ lysates, despite the expectation that it would be present in HT29 APC$^{+/-}$ (Figure 5.9). Instead, lower molecular weight bands were seen in HT29 and HT29 APC$^{+/-}$ lysates at approximately 100 kDa and 200 kDa, which are the expected sizes of the truncated proteins in HT29 (Cristofaro et al., 2015). The 200 kDa protein in the HT29 *APC*$^{+/-}$ lysate was of much lower intensity than in HT29 whereas the lower band at 100 kDa appeared stronger. This was unexpected, given that the more 5' APC mutation was corrected, meaning the lower 100 kDa band should have been the one to disappear in the edited cell line.

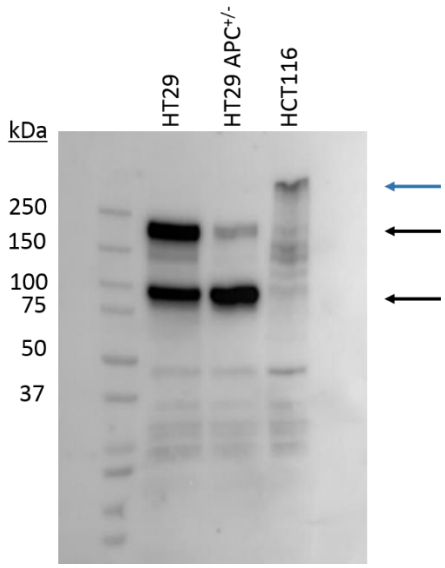

Figure 5.9. Expression of APC in CRC cell lines. Western blotting of cell lysates with an N-terminal antibody recognising APC. A long exposure was used to ensure any APC present in the lysates was seen thus there is considerable background. A band is seen in the HCT116 lane at the expected size of 312 kDa (blue arrow) but is absent in both the HT29 and HT20 APC$^{+/-}$ cell lysates. Bands at approximately 200 kDa and 100 kDa (black arrows) represent the truncated versions of APC found in HT29 cells.

Cell lysates were also blotted and probed with an antibody against the C-terminal region of APC. In this case, only those cell lines with full-length *APC* would be expected to give a signal (Figure 5.10). However, all cell lysates showed a band above 250 kDa irrespective of their *APC* status. Similarly all cell lysates gave a signal just above the 150 kDa band, which was not expected for any APC isoform. Previously, C-terminal antibodies that recognize the APC protein have been found to bind to an unrelated protein of approximately 150 kDa (Davies et al., 2007). This may be what is occurring here and, if so, this suggests that the C-terminal antibodies may be prone to non-specific binding.

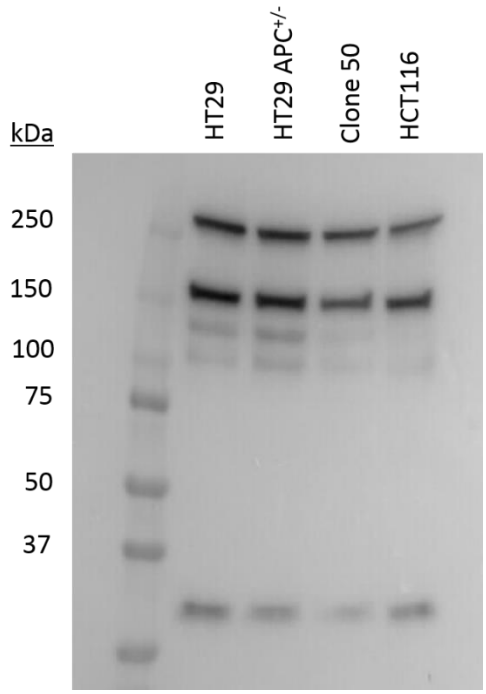

Figure 5.10. Western blotting of cell lysates with a C-terminal APC antibody. All cell lines gave strong signals for a band in excess of 250 kDa irrespective of their expected *APC* genomic status. Non-specific signal was also seen strongly in all lysates at about 150 kDa. Clone 50 was included as a possible *APC*-corrected cell line but was subsequently found to have suffered an eight amino acid in-frame deletion upstream of the targeted mutation.

5.3.8 Immunoprecipitation of proteins reveals specificity of APC antibodies.

To ascertain the specificity of the antibodies being used for probing for APC, N-terminal and C-terminal APC antibodies were used to immunoprecipitate the proteins they bind to by incubating the antibodies with HT29, HT29 *APC*$^{+/-}$ and HCT116 cell lysates. The precipitated

proteins were then separated by SDS-PAGE, transferred to PVDF membrane and probed with APC-specific antibodies. When the C-terminal antibody was used, no signal was obtained from the proteins precipitated with the N-terminal antibody (not shown). When the blot was probed with the N-terminal antibody, there was evidence of a large band in the HCT116 cell line lane precipitated by the C-terminal antibody (Figure 5.11). However, the same antibody failed to precipitate bands in either of the other two cell lines suggesting that the C-terminal bands for HT29 and HT29 *APC*+/- seen in Figure 5.10 above are non-specific. Thus while the C-terminal antibody can recognize full-length APC, it also has strong association with something other than APC and is therefore not a good antibody to use in our research.

It was also noted that there was only a faint band in the HCT116 cell line lane that was precipitated with N-terminal antibody, suggesting that the N-terminal antibody has reduced affinity for full-length APC (when compared to the strength of the band in the C-terminal precipitated HCT116). This also reflects the relative strengths of the bands seen in Figure 5.9 where whole lysates were stained with N-terminal antibody. There the full-length APC band in the HCT116 cell line was markedly less intense than the bands for truncated APC in the other lanes. Thus the N-terminal antibody we used may not be an ideal antibody for identifying full-length APC.

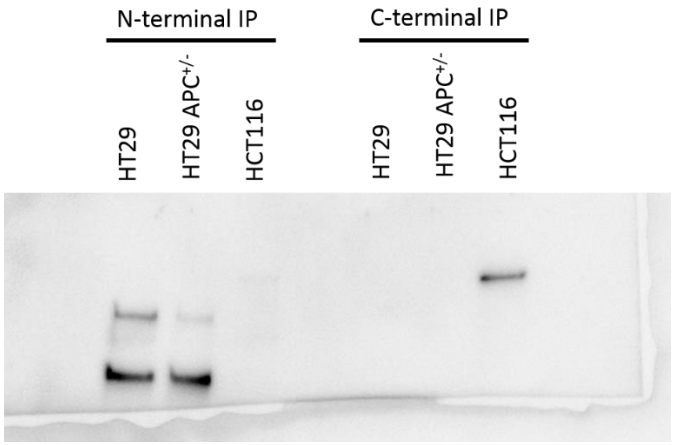

Figure 5.11. APC N-terminal probing of immunoprecipitated cell lysates. Proteins immunoprecipitated with either N-terminal or C-terminal antibodies were probed with the N-terminal antibody. Only HCT116 lysates show evidence of a single C-terminal-precipitated protein.

5.4 Discussion

The use of colonic epithelial cell lines to study environmental factors affecting colon carcinogenesis is well established. For example, HT29 and its derivative HT29/c1 are the predominant cell lines used in the study of ETBF infection (Sears, 2009). However, HT29 cells have only truncated forms of the APC protein due to mutation of the gene and we hypothesised that the absence of full length APC in HT29 cells may play a part in the marked sensitivity of HT29 cells to BFT toxin, a reason these cells are attractive to researchers studying ETBF in CRC. If loss of APC is indeed a factor in the response of colonic epithelial cells to environmental factors in the development of CRC, then HT29 cells are not ideal for the study of the earliest events in colorectal carcinogenesis.

We reasoned it would be intriguing to look at the same responses in the same cells with full length APC. To this end, we attempted to edit one allele of *APC* in HT29 cells using a CRISPR-Cas9n system and to characterise the resultant cell line. We successfully corrected the *c.2557G>T* mutation in *APC* by inducing single strand breaks close to each other on each DNA strand around the mutated base with a nickase enzyme, and then incorporating correct DNA sequence via homology-directed repair from template DNA inserted into the cell in a plasmid (Ran et al., 2013). The use of a nickase enzyme to create two nearby nicks on opposite strands, rather than a single double-strand break created by conventional Cas9, is thought to limit the chance of off-target effects, due to the reduced likelihood of both guide RNA recognition sequences annealing close to each other at non-specific sites (Ran et al., 2013).

The efficiency of the editing was low, with one successful clone being produced from 96 tested clones. This may be partly due to our use of separate Cas9n plasmids for each gRNA recognition site. These plasmids contain the same puromycin resistance gene for selection, potentially resulting in survival of cells containing one or both plasmids. Single strand nicks are readily fixed by repair processes in mammalian cells, and therefore the presence of only one guide RNA would result in Cas9n nicking one DNA strand, but this nick would be fixed by the cell in the absence of the second gRNA targeting Cas9 to create the second nick. Thus clones surviving antibiotic selection due to the presence of only one guide RNA encoding plasmid are unlikely to result in editing of *APC*, and this likely accounts for the high number of failed clones. An alternative strategy would be to include both guide RNA coding sequences in the same plasmid as has recently been done by others (Adikusuma et al., 2017).

We also encountered issues with the editing process where a clone (Clone 50) contained a 24-bp in-frame deletion directly upstream of the mutant base. This is likely to have resulted following successful nicking of the *APC* gene followed by incorporation of the deletion during the homology-directed repair process (Figure 5.2). Mistakes like these are common during CRISPR-Cas9 gene editing (van Overbeek et al., 2016) and therefore sequence analysis of the DNA around the edited bases is essential.

Our finding via Next Generation Sequencing that HT29 cells contain three copies of *APC* comprising two separate alleles, two copies of one allele with the *c.4666-4667insA* mutation and one copy of a second allele with the *c.2557G>T* mutation, while incidental to our project aims, is a noteworthy discovery. It fits with the previous finding that HT29 contains 3 copies of most of chromosome 5 (Abdel-Rahman et al., 2001) and will inform those working on *APC* in HT29 cells. Indeed with regard to the current project it is useful to know that the allele we attempted to correct represents only a third of the mutated *APC* alleles in HT29 cells, which may have relevance when considering the behaviour of successfully edited clones. Our finding that HT29 $APC^{+/-}$ cells have a markedly slower growth rate than the parent HT29 cells is in line with the study by Morin *et al* (Morin et al., 1996). This finding of a slower growth rate in cells expressing full length APC can be explained by the observation of a gradient of increasing APC expression levels in colonic epithelial cells undergoing crypt-to-villus differentiation, while highly proliferating cells at the bottom of the crypt express relatively little (Boman & Fields, 2013). *APC* mutation results in a shift in this gradient and a resulting dysregulation of cell proliferation up the crypt. The observed increase in the proportion of non-viable HT29 $APC^{+/-}$ cells (as measured by trypan blue staining) likely reflects the turnover of fully-differentiated cells occurring at the colonic epithelial surface.

The most striking property of the CRISPR-edited HT29 $APC^{+/-}$ cell line was its apparent resistance to the toxin-mediated loss of morphology that is widely reported in the HT29 cell line (Sears, 2009). Instead, after 6 hours of incubation with the concentrated toxin-containing ETBF supernatant, the HT29 $APC^{+/-}$ cell line showed no evidence of cell rounding, retaining morphology that was similar to that seen in a similar number of HCT116 cells (that are $APC^{+/+}$) (El-Bahrawy et al., 2004), treated with the same concentration of toxin and for the same length of time. Fluorescence microscopy confirmed that the resistance to morphological change upon exposure to ETBF supernatant exhibited by the edited HT29 $APC^{+/-}$ cell line was not due to any inability of BFT to cleave E-cadherin in these cells, suggesting that the mechanism by which HT29 cells round up following E-cadherin cleavage was prevented in the edited cells.

APC is a multi-functional protein and is known to interact with actin to stabilise the cytoskeleton (Moseley et al., 2007). Actin filaments are formed by the polymerisation of actin monomers in the cytoplasm. This process is started by specific proteins acting as actin nucleolators, an important one of which is APC. Actin filaments are polymerised by an interaction of APC and mDia1, a formin protein, with actin via the APC basic domain (Breitsprecher et al., 2012). APC is also key to maintaining the cytoskeleton through its binding to actin filaments via β-, and α-catenins (Narayan & Roy, 2003). Thus, while cells exposed to BFT have E-cadherin cleaved, the structure of the cells may be maintained in cells where APC can preserve the connections with the cytoskeleton, such as seen in BFT-treated HCT116 cells.

The truncated versions of APC present in HT29 cells both lack the C-terminal half of the protein which contains the basic domain required for actin filament assembly as well as catenin-binding sites. In particular, the *c.2557G>T* APC mutation that we have attempted to correct results in a protein with no catenin binding sites. Therefore HT29 cells may be more susceptible to rounding up upon BFT-mediated E-cadherin cleavage due to the inability of truncated APC to form actin filaments, and the presence of a full-length APC protein may restore this function.

In addition, APC forms a complex with EB1, via a binding site at the C-terminus of APC (Figure 5.1) and this APC/EB1 complex binds to the ends of microtubules which can then be coupled to the cell cortex (Mimori-Kiyosue et al., 2000). Reciprocal interaction between microtubules and the actin cytoskeleton has been shown to be necessary for the establishment and maintenance of cell shape (Akhmanova & Hoogenraad, 2015), and therefore the roles of the C-terminal half of APC in both of these interacting structures suggest the re-introduction of full-length APC to HT29 cells would likely have a stabilising effect on cell morphology (Figure 5.7) and actin cytoskeleton (Figure 5.8), even in the event of E-cadherin cleavage as seen in this study.

Cleavage of E-cadherin by BFT leads to release of membrane-associated β-catenin into the cytoplasm, where it can migrate to the nucleus, bind the transcription factor TCF4 and dysregulate Wnt-signalling target genes (Clevers, 2006). This promotes epithelial-mesenchymal transition (EMT) in the cell, a key step in colorectal carcinogenesis (Polyak & Weinberg, 2009). Where full-length APC exists in the cytoplasm, this excess β-catenin can be mopped up by the APC protein complex, phosphorylated, ubiquitinated and targeted for degradation (Clevers, 2006), thereby slowing the EMT process. In contrast, where full-length

APC is absent or partially absent (as in $APC^{Min+/-}$ mice), EMT and subsequent tumour development can occur rapidly following exposure to *B. fragilis* toxin (Rhee et al., 2009). Our finding of a slower growth rate of HT29 $APC^{+/-}$ cells, compared to the parent HT29 cells, also supports the assertion that APC slows down or reverses EMT within these cells, as the rate of proliferation is considerably decreased.

The proportion of non-viable cells was also considerably higher in the edited cell line than either HT29 or HCT116 cells. While the reason for increased cell death was not investigated here, Morin and colleagues observed the same phenomenon when full-length APC was induced and, by visualising condensed chromatin or micronuclei with Hoechst 33258 stain in the floating cells, concluded that cell death was due to apoptosis (Morin et al., 1996).

The failure to identify APC protein expression in the edited HT29 $APC^{+/-}$ cell line is concerning. The change in cell growth rate and the resistance of the edited cells to morphological change upon exposure to ETBF supernatant were predicted outcomes of successful expression of full-length APC generated by the editing process. Induced expression of APC in HT29 led to a substantial decrease in the growth rate of the cells (Morin et al., 1996) and HCT116 cells are resistant to morphological change despite loss of E-cadherin. Accordingly, showing expression of full-length APC in the edited cell line will be critical to proving our hypothesis that BFT-mediated morphological change in HT29 cells is enabled by the lack of full-length APC, and the resistance to this shown by the edited cells is not due to an as yet unidentified factor. Despite the choice of antibodies in this chapter being based on a review of previously published research, there does seem to be a question mark over the reliability of both C-terminal and N-terminal APC antibodies (Davies et al., 2007). This lack of availability of APC antibodies that can consistently immunoprecipitate and immunoblot effectively may limit progress. One additional way to investigate this could be to investigate expression of Wnt-signalling target genes such as *MYC*, which encodes the c-myc transcription factor, based on an early study that reports expression of full-length APC should result in a decrease in c-myc (T. C. He et al., 1998).

Alternatively, or additionally, the inability to detect wild-type APC in the edited cell line could be due to the initial clone containing unedited HT29 cells, possibly protected from puromycin selection by surrounding edited cells. Upon removal of puromycin, the faster growth of these cells may have led to them be over-represented in subsequent culture and therefore making wild-type APC harder to find. Both next generation and Sanger sequencing of early cultures of

Clone 30 indicate successful editing of the *APC* gene and the growth and behaviour of the cell line seem to indicate that full-length APC is present. Later failure to show full-length APC protein may be due to poor antibodies, the presence of unedited HT29 cells or a combination of both.

In conclusion, the editing of the *c.2557G>T* mutation in *APC* in HT29 cells resulted in a cell line that is apparently resistant to the BFT-mediated loss of cell morphology that is seen in toxin-treated HT29 cells. While our results suggest this may relate to the association of APC to E-cadherin via actin and the catenins, further experiments are necessary to prove this hypothesis. Nevertheless, the altered response of the edited cells to BFT-enriched supernatant indicates this newly derived cell line will be a useful tool in increasing our understanding of the role of BFT in colorectal carcinogenesis.

Chapter 6 Discussion and future directions

6.1 General Discussion

Early-onset CRC incidence is increasing worldwide despite a reduction in overall incidence of CRC, and New Zealand has among the steepest rises in incidence among young adults (Siegel et al., 2019). It has been postulated that early-onset CRC constitutes a separate subset of CRC (Silla et al., 2014), therefore a greater understanding about the aetiology of these cancers, including factors that potentially influence early-stage carcinogenesis, is paramount to treatment, detection and prevention strategies.

To this end, the research detailed in this thesis comprised three distinct yet related studies. Firstly, the *CDH1* gene was sequenced and the expression of its product, E-cadherin, was investigated by immunohistochemistry in a cohort of colorectal signet ring cell carcinomas from Pakistanis aged 40 years or under. The rationale for this was based on the similarities between colorectal SRCCs and SRCCs in hereditary diffuse gastric cancer, and the strong link between mutation of *CDH1* and HDGCs (Guilford et al., 1998). While colorectal SRCCs make up approximately 1% of all primary CRCs (Kang et al., 2005), they are frequently found in younger patients with one study of 29 SRCC patients having 18 (62%) aged under 50 years old (Anthony et al., 1996).

Secondly, this study was expanded to a local New Zealand cohort of early-onset colorectal tumours, primarily adenocarcinomas. In addition to *CDH1* we also sequenced *APC*, a gene mutated in 60-80% of colorectal cancers but with reportedly lower levels in younger patients (Willauer et al., 2019). The early-onset CRC cohort contained only one SRCC and one adenosquamous carcinoma while the rest were conventional adenocarcinomas, suggestive of a lower level of *CDH1* mutations and E-cadherin loss. Due to its frequent occurrence in CRC, *APC* mutation was likely to be present more frequently than *CDH1* mutation. Moreover, the reported rate of *APC* mutation in EOCRC compared to older colorectal tumours made this an interesting gene to investigate.

Finally, a mutated *APC* gene in a commonly used colorectal cell line, HT29, was edited using a CRISPR/Cas9n approach, to produce a cell line which expresses full length APC protein. The rationale for this was the growing awareness that early-life exposures potentially influence colon carcinogenesis (Siegel et al., 2019). To that end, we exposed the wild-type and

CRISPR/Cas9n edited cells to a bacterial toxin linked to the development of pre-cancerous lesions in the colon (Purcell et al., 2017), to better understand the role of APC in early colorectal carcinogenesis.

6.1.1 *CDH1* mutations in SRCC

The finding, described in Chapter 1, of multiple *CDH1* mutations in the colorectal SRCC samples from Pakistan (Aitchison et al., 2020), was not surprising due to the histological similarities between colorectal SRCC and HDGC, where *CDH1* mutations are the predominant cause (Guilford et al., 1998). The loss of functional E-cadherin leads to loss of cell to cell contact and, coupled with the accumulation of large volumes of intracellular mucin, results in the classic signet-ring morphology where the nucleus is pushed to the edge of a rounded cell (Gopalan et al., 2011).

With the samples coming from a region of the world where consanguineous marriages are common, the high rates of early-onset SRCCs observed may have suggested an autosomal recessive disorder involving *CDH1*. However, the presence of wild type alleles alongside the mutant forms suggested this was not the case.

There was also no evidence of a founder mutation effect in the cohort, as no common mutations were seen in this cohort. This absence of a single founder mutation effect is also notable amongst New Zealand Māori, with high rates of HDGC caused by inherited *CDH1* mutations, where many different mutations are found (Hakkaart et al., 2019), whereas a founder *CDH1* mutation has been identified in several Newfoundland families with HDGC (Kaurah et al., 2007). The finding of multiple different *CDH1* mutations in a population, if germline, as in the New Zealand Māori cohort, would suggest aberrant E-cadherin has offered a selective advantage at some point. The nature of any selective advantage is unknown. E-cadherin is however a target for adhesion of *Candida albicans* (Phan et al., 2007), *Listeria monocytogenes* (Mengaud et al., 1996) and *Streptococcus pneumonia* (Anderton et al., 2007). Thus aberrant E-cadherin protein structure and/or expression may offer some resistance to adhesion of these (and potentially other) pathogenic microorganisms.

Despite the lack of a founder effect or autosomal recessive inheritance, the possibility that some of the mutations found in the colorectal SRCC cohort were inherited was intriguing. However, the absence of family history for any of the Pakistani patients that could have provided evidence

of germline mutation in some of the cases precluded further investigation of this aspect of the study. Additionally, the lack of clinical follow-up data prevented assessment of any prognostic significance of loss of E-cadherin (Wang et al., 2016).

It was noted that there was not a complete concordance of *CDH1* mutation status with E-cadherin expression in the SRCC samples. Tumours lacking *CDH1* mutations, but also with weak or absent E-cadherin expression, represent alternative mechanisms of E-cadherin silencing, such as allelic loss or epigenetic silencing. Epigenetic silencing in the form of promoter methylation, where methyl binding proteins such as MeCP2 bind to methylated cytosine residues in the promoter preventing transcription factors from binding to the DNA, is a likely possibility as has been seen in HDGC either on its own or as a second-hit following *CDH1* mutation (Grady et al., 2000). *CDH1* promoter methylation in colorectal SRCCs has previously been shown in colorectal cancers (Darwanto et al., 2003) however methylation analysis of the Pakistani samples was not possible due to the poor yield and quality of DNA.

Allelic loss at the *CDH1* locus is another possibility for the two tumours showing a majority of alleles to be mutant. Loss of heterozygosity has been reported as a common feature of colorectal SRCCs (Kakar & Smyrk, 2005) and may account for the lack of wild type *CDH1* sequences for both the E-cadherin negative staining sample with a predicted damaging p.G839D mutation, and the weakly E-cadherin positive sample with a predicted benign p.A446V mutation. Unfortunately, as with methylation analyses, the yield and quality of the DNA recovered from the FFPE samples precluded an investigation of LOH in these samples.

Similarly, PCR investigations of microsatellite instability would not have been possible in these samples. However, immunohistological assessment of MSI was possible by staining for the mismatch repair proteins MSH6 and PMS2. While some previous reports have suggested MSI may be a feature of SRCCs, there is no consensus on the association of MSI with SRCC (Nam et al., 2018; Ogino et al., 2006; Rosty et al., 2014). Interestingly, the finding of only two samples in this study having MSI, of which only one was a SRCC, the other being an adenocarcinoma with a signet ring component less than 50%, adds to the evidence of MSI not being a feature of SRCC (Nam et al., 2018). Furthermore, the two samples that lacked detectable PMS2 expression both showed positive, although weak, E-cadherin staining, suggesting that MSI and E-cadherin loss may represent different pathways of colorectal SRCC development.

The different levels of E-cadherin staining may also reflect differences in the phenotype associated with the particular mutation. A recent study looking at different *CDH1* mutations suggested the location and type of *CDH1* mutation may play a role in the phenotype displayed, with differences seen between patients with cleft lip or palate and those with HDGC (Selvanathan et al., 2020).

Overall, the results detailed in Chapter 3 reinforce that E-cadherin loss is a key feature in colorectal SRCCs, but suggest that this loss can be due to a number of different factors including, in some but not all cases, *CDH1* mutation. Further insight into potential causes, however, was precluded by two significant issues which were the quality of the FFPE samples and the lack of any data relating to the clinical outcome of each case, or to their individual family history of CRC.

6.1.2 *CDH1* and *APC* mutation in early-onset CRC

Finding multiple *CDH1* mutations in colorectal SRCC tumours from young people, we hypothesized that *CDH1* mutation may contribute to the early-onset of colorectal cancer in young patients in New Zealand. However, this hypothesis did not stand up to testing as only one sample from our cohort was found to have a mutation and this was predicted to be benign. In addition E-cadherin staining was positive in all but one sample. Indeed the sample negative for E-cadherin was the only SRCC sample in the New Zealand cohort, reinforcing that loss of E-cadherin is a feature of colorectal SRCCs but not colorectal adenocarcinomas. Interestingly, this SRCC sample had no *CDH1* mutation, was microsatellite stable and had no family history of cancers suggesting the lack of E-cadherin could be due to methylation of the *CDH1* promoter (Darwanto et al., 2003). This remains to be investigated.

While *CDH1* mutation was not found to be a feature of early-onset CRC in this cohort, it should be noted that the ethnic make-up of the cohort was almost exclusively European except for a single Filipino patient. The lack of Māori in the cohort prevents the investigation of *CDH1* mutation as a cause of early-onset CRC in that group which gains significance in the light of *CDH1* mutation as a cause of HDGC in Māori (Guilford et al., 1998). This may be remedied by studying a larger cohort including samples from regions of New Zealand where Māori populations are higher to create a cohort more representative of the New Zealand population as a whole.

An earlier study of 50 New Zealand colorectal cancer patients under the age of 25 years reported five tumours (10%) as SRCCs (Plunkett et al., 2014), whereas only one SRCC was found in the cohort detailed in chapter 4. This difference may, in part, reflect the ages of the two groups (median age 23.6 and 44 years, respectively). A higher incidence of SRCC in a younger patient cohort may reflect the observation that the younger the patient, the poorer the outcome of CRC (Poles et al., 2016), as SRCC generally confers a poor prognosis (Barresi et al., 2017). A further distinction of CRC in patients of a very young age is a reduced *APC* mutation rate only being found in the under 30 year age bracket, in a recent study looking at molecular differences in CRC at different age brackets (Willauer et al., 2019). These contrasts between age groups suggest the assertion, that early-onset CRCs are a subset of CRCs (Silla et al., 2014), may be particularly relevant for those patients under 30 years.

The finding of *APC* mutations in the New Zealand cohort was not in itself surprising, but the rate of mutations (72%) was higher than previously reported in early-onset CRCs (Kothari et al., 2016; Willauer et al., 2019). This is likely to be due to the sequencing strategy used. Many studies have focused on the mutation cluster region as this is where the bulk of mutations are located, particularly in sporadic CRCs (Fearon, 2011). However, our strategy looked, and found, mutations at many locations upstream of the mutation cluster region. If only the mutations found in the mutation cluster region were considered then our rate of *APC* mutation, 32% within the mutation cluster region and 56% within the final exon, would have been similar to previous reports. Two different cohorts investigated in one study had *APC* mutation rates for under 50 year olds of 41.6% and 65.2% respectively, with each cohort being studied using different protocols (Willauer et al., 2019). Accordingly, our high rate of *APC* mutations in upstream regions of the gene could be due to other studies confining their search. However, our finding also raises the possibility of an association between upstream *APC* mutations and early-onset CRCs. In FAP, germline *APC* mutations in the 5' portion of the gene are associated with an attenuated form of the disease, whereas mutations within the mutation cluster led to conventional FAP, indicating phenotypic differences depending on the site of an *APC* mutation (Christie et al., 2013). These days modern sequencing techniques allow the investigation of large regions of DNA relatively cheaply. Accordingly, sequencing of the whole *APC* gene in future studies would ensure *APC* mutations were not missed, and expand the understanding of the full scale of *APC* mutations in early-onset CRCs.

While the majority of *APC* mutations found in our cohort were sporadic and resulted in truncated proteins as expected, the discovery of a p.E1317Q missense mutation was of

particular interest. This mutation has been described before and appears to have incomplete penetrance of CRC (Frayling et al., 1998; White et al., 1996), suggesting environmental factors may determine whether individuals with p.E1317Q develop tumours. The patient in this study with the p.E1317Q mutation had a great aunt with bowel cancer, both parents had non-GI cancers and the patient was subsequently referred for testing for HNPCC. However, testing for FAP was not requested, presumably based on the absence of polyps. The absence of polyps in this case may be due to the mutation being a missense mutation, rather than the more common truncating mutation observed in APC. As FAP is caused by mutations in the *APC* gene and codon 1317 is situated within the mutation cluster region, testing would have uncovered the mutation.

6.1.3 Gene editing of mutant *APC* in HT29 cells

The *APC* gene is mutated in most colorectal cancers (Kinzler & Vogelstein, 1996) and mutation is thought to be an early event in the adenoma-carcinoma pathway of colorectal carcinogenesis (Fearon & Vogelstein, 1990). The finding, described in Chapter 4, that 72% of early-onset CRC patients carried a mutation in *APC*, underlies the importance of understanding how the APC protein prevents carcinogenesis, and how mutation of *APC* is key to the cascade of changes that result in CRC.

The HT29 colon cancer cell line has two truncated isoforms of APC and exhibits a strong response to the toxigenic bacterium ETBF, resulting in cell rounding. This has been attributed to the toxin-mediated loss of E-cadherin (Sears, 2009), but our observation that introducing full length APC protein to HT29 cells allowed the cells to maintain their structure after 6 hours of exposure to ETBF supernatant, suggests a role for APC in preserving cell structure. Fluorescent immunohistochemistry showed that E-cadherin was cleaved by the toxin in the edited cell line just as in the parent HT29 cells, so the maintenance of cell shape was due to some internal process involving APC and not resistance to E-cadherin cleavage.

While the behaviour of the edited cell line suggested that full length APC was being expressed, and the sequence of the edited *APC* gene suggested successful correction of the mutated base, it was not possible to visualize full length APC. This may be due to non-specificity of APC antibodies. While the antibodies were chosen based on the literature they failed to work consistently in this study. The N-terminal antibody failed to consistently immunoprecipitate

and immunoblot (Figure 5.11), while the C-terminal APC antibody consistently picked up non-specific protein, as reported elsewhere (Davies et al., 2007).

6.2 Future Directions

The increase in recent years in the number of people under 50 years of age in New Zealand (Gandhi et al., 2017) and around the world (Vuik et al., 2019) being diagnosed with sporadic CRC, while the rates of CRC in older patients decline, is an as yet unexplained phenomenon. The aim of this and future research is to better understand whether early onset CRC constitutes a subset with a unique biology, or whether it presents a continuum of older onset CRC, perhaps exacerbated by increasing and younger exposure to environmental and/or modern lifestyles (J. Keenan et al., 2017; Richardson et al., 2016). The worldwide increase in early onset CRC, coupled with the finding that a family history of CRC is not necessarily a risk factor for this disease, increasingly supports the idea that modifiable risk factors such as diet, obesity and lack of exercise are likely to have a role (Connell et al., 2017), and this is reinforced by studies on people who migrate from low- to high-risk areas of the world (Flood et al., 2000).

Accordingly, considering lifestyle factors in conjunction with clinical factors and genetic analyses may help shed light on the reasons for the increase in early onset CRC. For example, patients with tumours exhibiting *APC* mutations have been found to have higher alcohol consumption, and red and processed meat consumption, than patients with *APC* wild type tumours (Gay et al., 2012). While the research described in this thesis compared *APC* mutation status with clinical data, such as tumour location and age, further research with access to lifestyle factors including diet, exercise, smoking and alcohol consumption, could provide important insights into environmental factors affecting cancer risk.

An important factor not considered in this work, with regard to *CDH1* and *APC*, is promoter methylation as a form of epigenetic silencing (Darwanto et al., 2003; Esteller et al., 2000). While the quality of the samples in Chapter 3 precluded study of methylation, the samples studied in Chapter 4 are suitable for methylation analysis and only time constraints prevented their inclusion in this thesis. Gene expression analysis of *CDH1, APC* and downstream targets, such as *MYC* and *SMO* in these tumours, will complement the methylation studies. In addition to the direct correlation of promoter methylation to gene expression, methylation analysis of

these and other genes may also link with lifestyle and the other environmental factors discussed above as a future area of research (Gay et al., 2012; Wong et al., 2007).

The choice to investigate only *CDH1* in the Pakistani cohort, and *CDH1* and *APC* in the NZ cohort, rather than looking more widely throughout the genome was taken for a number of reasons. Firstly, there were solid hypotheses around looking at *CDH1* and *APC* within these cohorts based on previous work by other groups. Secondly, there was an existing protocol for *CDH1* sequencing (Hakkaart et al., 2019), which it was possible to use and adapt to sequence the *APC* gene as well. Thirdly, by using unique combinations of indexed primers for each sample, the number of flow cells run was kept to a minimum allowing the studies to be carried out at low cost. Despite the reducing costs of whole genome sequencing, and potential additional information that might have been obtained, that approach would have been prohibitively expensive in this case.

The correction of one allele of *APC* by gene editing resulted in a colon cancer cell line (HT-29) that showed resistance to morphological change when incubated with a *B. fragilis* toxin-enriched supernatant, suggestive of the expression of full-length APC. However, the failure to detect APC protein in the edited cells leaves questions as to whether the stability of cell morphology is indeed due to the proposed mechanisms. Showing that full-length APC is indeed expressed will be an important first step in using these cells as a model of bacterially-induced carcinogenesis. Our finding that APC antibodies may not be reliable, reflecting previous findings (Davies et al., 2007), means that finding a robust APC antibody will be a priority for the next step in this research. An alternative antibody, FE-9, raised against the amino terminal of APC, was used successfully to detect full-length APC in HT29 cells containing a plasmid with an inducible APC gene (Morin et al., 1996). Additionally, sequencing of mRNA from the edited cell line should show no *APC* transcripts with the c.2557C>G mutation if the *APC* gene has been successfully edited. A further sequencing approach would be to use the long reads possible with Nanopore sequencing to show single DNA amplicons with both wildtype c.2557 and wildtype c.4666 which would provide further proof that the editing has been successful. It was noted that in their paper discussing inducible expression of full-length APC in HT29 cells, Morin and colleagues mention that their attempts to constitutively express APC in HT29 cells were unsuccessful (Morin et al., 1996). This, combined with our inability to detect full-length APC in the edited cell line, suggests that HT29 cells may be unable to express APC, or that cells expressing APC are swiftly outgrown by cells with only truncated forms of APC. If this

were the case it might be expected that wildtype APC expression is being silenced, perhaps by some epigenetic mechanism. This is an area of ongoing investigation.

APC has a key role in control of the Wnt/β-catenin signalling pathway, forming a complex with axin and glycogen synthase kinase 3β, to phosphorylate β-catenin and target it for ubiquitination and subsequent lysosomal degradation (Nathke, 2006). Expression of full-length APC in HT29 cells should therefore result in a reduction of Wnt-signalling target gene expression, particularly *MYC*, as seen when an inducible form of APC is expressed in HT29 cells (T. C. He et al., 1998). Analysis of RNA expression will be an important future task, not just for *MYC* but also other genes. We would expect to see expression changes in the presence of full-length APC, such as reduced *SMO,* a mediator of DNA damage through generation of reactive oxygen species, that is frequently upregulated in CRC (Goodwin et al., 2011), and proteins in the Notch signalling pathway, which display crosstalk with Wnt signalling in the control of cell proliferation (Vinson et al., 2016).

The epithelium of the colon consists of a polarised monolayer of cells that form along crypts. We have proposed using the HT29 $APC^{+/-}$ cells as a model of intestinal epithelial cells to study the role of ETBF in colorectal carcinogenesis. Colorectal neoplastic cells develop from differentiated epithelial cells before spreading along the crypt in a top-down morphogenetic process (Shih et al., 2001). However, HT29 cells, and the edited HT29 $APC^{+/-}$ cells are undifferentiated epithelial cells. HT29 cells can be differentiated by growth in glucose-free medium, where glucose is replaced as a carbohydrate source by galactose (Huet et al., 1987). These HT29 derivative cells, known as HT29/c1, grow as polarised differentiated cells. If the HT29 $APC^{+/-}$ cells can be similarly differentiated by growth in glucose-free media, the effects of ETBF on polarised monolayers containing full length APC can be studied.

An alternative approach to CRISPR-editing of *APC* in HT29 cells that was considered involved using the approach of Huet and colleagues mentioned above to differentiate the HCT116 colonic epithelial cell line that has full-length APC. Despite being able to reproduce the published differentiation in HT29 cells using this approach, we were unable to obtain the same results with HCT116 cells. A second approach using sodium butyrate to differentiate HCT116 cells (Fung et al., 2009) also failed, despite verifying the process in HT29 cells. These failures to differentiate HCT116 cells by methods successful in HT29 cells emphasise the differences between these CRC cell lines, and that understanding the mechanisms underlying the sensitivity of HT29 cells to BFT is critical to understanding any role of BFT in inducing

carcinogenesis. This gains significance when considering almost all cellular work on BFT-induced carcinogenesis has been carried out on HT29 cells.

Another alternative approach to studying the role of BFT in the initiation of CRC would be to grow primary colonic epithelial cells isolated from normal-looking mucosa of patients, and to assess the morphological, gene expression and protein responses following exposure to BFT. BFT exposure of primary colonic tumour cells has been studied previously (Sanfilippo et al., 1998) and adaptation of this protocol to study normal colonic cells as a model of normal epithelium could be a useful tool in investigating CRC initiation.

In summary, disruption of E-cadherin was found to occur frequently in colorectal SRCC, whether by *CDH1* mutation or by other mechanisms, but this was not the case with early-onset CRC, where all non-SRCC tumours expressed E-cadherin. Conversely, *APC* mutation is widespread in early-onset CRC, primarily sporadic but can be germline, even in a cohort where known familial cases have been excluded, suggesting that criteria for testing for familial syndromes may exclude some families.

Rates of *APC* mutation were higher than other reports of early-onset CRC. Whereas this may be due to methodological difference, future strategies should look at the whole *APC* gene rather than just the mutation cluster region or other hotspots. This gains further significance with our novel observation that establishment of a derivative cell line of HT29 expressing full-length APC provided evidence that this protein appears to have an underappreciated role in stabilising cellular morphology. Accordingly, the derivative cell line is likely to prove a useful tool for studying many aspects of colorectal carcinogenesis.

References

Abdel-Rahman, W. M., Katsura, K., Rens, W., Gorman, P. A., Sheer, D., Bicknell, D., Bodmer, W. F., Arends, M. J., Wyllie, A. H., & Edwards, P. A. (2001). Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement. *Proc Natl Acad Sci U S A, 98*(5), 2538-2543. doi:10.1073/pnas.041603298

Aceto, G., Curia, M. C., Veschi, S., De Lellis, L., Mammarella, S., Catalano, T., Stuppia, L., Palka, G., Valanzano, R., Tonelli, F., Casale, V., Stigliano, V., Cetta, F., Battista, P., Mariani-Costantini, R., & Cama, A. (2005). Mutations of APC and MYH in unrelated Italian patients with adenomatous polyposis coli. *Hum Mutat, 26*(4), 394. doi:10.1002/humu.9370

Adikusuma, F., Pfitzner, C., & Thomas, P. Q. (2017). Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression. *PLoS One, 12*(12), e0187236. doi:10.1371/journal.pone.0187236

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet, Chapter 7*, Unit7 20. doi:10.1002/0471142905.hg0720s76

Aitchison, A., Hakkaart, C., Whitehead, M., Khan, S., Siddique, S., Ahmed, R., Frizelle, F. A., & Keenan, J. I. (2020). CDH1 gene mutation in early-onset, colorectal signet-ring cell carcinoma. *Pathol Res Pract, 216*(5), 152912. doi:10.1016/j.prp.2020.152912

Akhmanova, A., & Hoogenraad, C. C. (2015). Microtubule minus-end-targeting proteins. *Curr Biol, 25*(4), R162-171. doi:10.1016/j.cub.2014.12.027

Al-Sohaily, S., Biankin, A., Leong, R., Kohonen-Corish, M., & Warusavitarne, J. (2012). Molecular pathways in colorectal cancer. *J Gastroenterol Hepatol, 27*(9), 1423-1431. doi:10.1111/j.1440-1746.2012.07200.x

Albuquerque, C., Breukel, C., van der Luijt, R., Fidalgo, P., Lage, P., Slors, F. J., Leitao, C. N., Fodde, R., & Smits, R. (2002). The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. *Hum Mol Genet, 11*(13), 1549-1560. doi:10.1093/hmg/11.13.1549

Amini, A. Q., Samo, K. A., & Memon, A. S. (2013). Colorectal cancer in younger population: our experience. *J Pak Med Assoc, 63*(10), 1275-1277. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24392559

Anderton, J. M., Rajam, G., Romero-Steiner, S., Summer, S., Kowalczyk, A. P., Carlone, G. M., Sampson, J. S., & Ades, E. W. (2007). E-cadherin is a receptor for the common protein pneumococcal surface adhesin A (PsaA) of Streptococcus pneumoniae. *Microb Pathog, 42*(5-6), 225-236. doi:10.1016/j.micpath.2007.02.003

Andriamihaja, M., Lan, A., Beaumont, M., Audebert, M., Wong, X., Yamada, K., Yin, Y., Tome, D., Carrasco-Pozo, C., Gotteland, M., Kong, X., & Blachier, F. (2015). The deleterious metabolic and genotoxic effects of the bacterial metabolite p-cresol on colonic epithelial cells. *Free Radic Biol Med, 85*, 219-227. doi:10.1016/j.freeradbiomed.2015.04.004

Anthony, T., George, R., Rodriguez-Bigas, M., & Petrelli, N. J. (1996). Primary signet-ring cell carcinoma of the colon and rectum. *Ann Surg Oncol, 3*(4), 344-348. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8790846

Armstrong, H., Bording-Jorgensen, M., Dijk, S., & Wine, E. (2018). The Complex Interplay between Chronic Inflammation, the Microbiome, and Cancer: Understanding Disease Progression and What We Can Do to Prevent It. *Cancers (Basel), 10*(3). doi:10.3390/cancers10030083

Aune, D., Lau, R., Chan, D. S., Vieira, R., Greenwood, D. C., Kampman, E., & Norat, T. (2011). Nonlinear reduction in risk for colorectal cancer by fruit and vegetable intake based on meta-analysis of prospective studies. *Gastroenterology, 141*(1), 106-118. doi:10.1053/j.gastro.2011.04.013

Bailey, C. E., Hu, C. Y., You, Y. N., Bednarski, B. K., Rodriguez-Bigas, M. A., Skibber, J. M., Cantor, S. B., & Chang, G. J. (2015). Increasing disparities in the age-related incidences of colon and rectal

cancers in the United States, 1975-2010. *JAMA Surg, 150*(1), 17-22. doi:10.1001/jamasurg.2014.1756
Ballester, V., Rashtak, S., & Boardman, L. (2016). Clinical and molecular features of young-onset colorectal cancer. *World J Gastroenterol, 22*(5), 1736-1744. doi:10.3748/wjg.v22.i5.1736
Bardou, M., Barkun, A. N., & Martel, M. (2013). Obesity and colorectal cancer. *Gut, 62*(6), 933-947. doi:10.1136/gutjnl-2013-304701
Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science, 315*(5819), 1709-1712. doi:10.1126/science.1138140
Barresi, V., Reggiani Bonetti, L., Ieni, A., Caruso, R. A., & Tuccari, G. (2017). Poorly Differentiated Clusters: Clinical Impact in Colorectal Cancer. *Clin Colorectal Cancer, 16*(1), 9-15. doi:10.1016/j.clcc.2016.06.002
Barrow, T. M., Klett, H., Toth, R., Bohm, J., Gigic, B., Habermann, N., Scherer, D., Schrotz-King, P., Skender, S., Abbenhardt-Martin, C., Zielske, L., Schneider, M., Ulrich, A., Schirmacher, P., Herpel, E., Brenner, H., Busch, H., Boerries, M., Ulrich, C. M., & Michels, K. B. (2017). Smoking is associated with hypermethylation of the APC 1A promoter in colorectal cancer: the ColoCare Study. *J Pathol, 243*(3), 366-375. doi:10.1002/path.4955
Beaugerie, L., & Itzkowitz, S. H. (2015). Cancers complicating inflammatory bowel disease. *N Engl J Med, 372*(15), 1441-1452. doi:10.1056/NEJMra1403718
Becker, K. F., Atkinson, M. J., Reich, U., Becker, I., Nekarda, H., Siewert, J. R., & Hofler, H. (1994). E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Res, 54*(14), 3845-3852. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8033105
Bellis, J., Duluc, I., Romagnolo, B., Perret, C., Faux, M. C., Dujardin, D., Formstone, C., Lightowler, S., Ramsay, R. G., Freund, J. N., & De Mey, J. R. (2012). The tumor suppressor Apc controls planar cell polarities central to gut homeostasis. *J Cell Biol, 198*(3), 331-341. doi:10.1083/jcb.201204086
Bettington, M., Walker, N., Clouston, A., Brown, I., Leggett, B., & Whitehall, V. (2013). The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology, 62*(3), 367-386. doi:10.1111/his.12055
Bhurgri, Y., Khan, T., Kayani, N., Ahmad, R., Usman, A., Bhurgri, A., Bashir, I., Hasan, S. H., & Zaidi, S. (2011). Incidence and current trends of colorectal malignancies in an unscreened, low risk Pakistan population. *Asian Pac J Cancer Prev, 12*(3), 703-708. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21627368
Bingham, S. A., Hughes, R., & Cross, A. J. (2002). Effect of white versus red meat on endogenous N-nitrosation in the human colon and further evidence of a dose response. *J Nutr, 132*(11 Suppl), 3522S-3525S. doi:10.1093/jn/132.11.3522S
Bisgaard, M. L., Fenger, K., Bulow, S., Niebuhr, E., & Mohr, J. (1994). Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. *Hum Mutat, 3*(2), 121-125. doi:10.1002/humu.1380030206
Bleyer, A. (2009). CAUTION! Consider cancer: common symptoms and signs for early detection of cancer in young adults. *Semin Oncol, 36*(3), 207-212. doi:10.1053/j.seminoncol.2009.03.004
Boland, C. R., & Goel, A. (2010). Microsatellite instability in colorectal cancer. *Gastroenterology, 138*(6), 2073-2087 e2073. doi:10.1053/j.gastro.2009.12.064
Boland, C. R., Koi, M., Chang, D. K., & Carethers, J. M. (2008). The biochemical basis of microsatellite instability and abnormal immunohistochemistry and clinical behavior in Lynch syndrome: from bench to bedside. *Fam Cancer, 7*(1), 41-52. doi:10.1007/s10689-007-9145-9
Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170
Boman, B. M., & Fields, J. Z. (2013). An APC:WNT Counter-Current-Like Mechanism Regulates Cell Division Along the Human Colonic Crypt Axis: A Mechanism That Explains How APC

Mutations Induce Proliferative Abnormalities That Drive Colon Cancer Development. *Front Oncol, 3*, 244. doi:10.3389/fonc.2013.00244

Borger, M. E., Gosens, M. J., Jeuken, J. W., van Kempen, L. C., van de Velde, C. J., van Krieken, J. H., & Nagtegaal, I. D. (2007). Signet ring cell differentiation in mucinous colorectal carcinoma. *J Pathol, 212*(3), 278-286. doi:10.1002/path.2181

Borrell, B. (2010). How accurate are cancer cell lines? *Nature, 463*(7283), 858. doi:10.1038/463858a

Bosman, F. T., World Health, O., & International Agency for Research on, C. (2010). *WHO classification of tumours of the digestive system* (4th ed. ed.). Lyon: IARC Press.

Botteri, E., Iodice, S., Bagnardi, V., Raimondi, S., Lowenfels, A. B., & Maisonneuve, P. (2008). Smoking and colorectal cancer: a meta-analysis. *JAMA, 300*(23), 2765-2778. doi:10.1001/jama.2008.839

Brattain, M. G., Fine, W. D., Khaled, F. M., Thompson, J., & Brattain, D. E. (1981). Heterogeneity of malignant cells from a human colonic carcinoma. *Cancer Res, 41*(5), 1751-1756. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7214343

Breitsprecher, D., Jaiswal, R., Bombardier, J. P., Gould, C. J., Gelles, J., & Goode, B. L. (2012). Rocket launcher mechanism of collaborative actin assembly defined by single-molecule imaging. *Science, 336*(6085), 1164-1168. doi:10.1126/science.1218062

Briske-Anderson, M. J., Finley, J. W., & Newman, S. M. (1997). The influence of culture time and passage number on the morphological and physiological development of Caco-2 cells. *Proc Soc Exp Biol Med, 214*(3), 248-257. doi:10.3181/00379727-214-44093

Buc, E., Dubois, D., Sauvanet, P., Raisch, J., Delmas, J., Darfeuille-Michaud, A., Pezet, D., & Bonnet, R. (2013). High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. *PLoS One, 8*(2), e56964. doi:10.1371/journal.pone.0056964

Bufill, J. A. (1990). Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med, 113*(10), 779-788. doi:10.7326/0003-4819-113-10-779

Burkitt, D. P., Walker, A. R., & Painter, N. S. (1974). Dietary fiber and disease. *JAMA, 229*(8), 1068-1074. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4407955

Cani, P. D., Osto, M., Geurts, L., & Everard, A. (2012). Involvement of gut microbiota in the development of low-grade inflammation and type 2 diabetes associated with obesity. *Gut Microbes, 3*(4), 279-288. doi:10.4161/gmic.19625

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R. A., & Holt, R. A. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res, 22*(2), 299-306. doi:10.1101/gr.126516.111

Chae, A., Aitchison, A., Day, A.S., Keenan, J.I. (2017). Bovine colostrum demonstrates anti-inflammatory and antibacterial activity in in vitro models of intestinal inflammation and infection. *Journal of Functional Foods, 28*, 293-298. doi:https://doi.org/10.1016/j.jff.2016.11.016.

Chan, D. S., Lau, R., Aune, D., Vieira, R., Greenwood, D. C., Kampman, E., & Norat, T. (2011). Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. *PLoS One, 6*(6), e20456. doi:10.1371/journal.pone.0020456

Chandra, S. H., Wacker, I., Appelt, U. K., Behrens, J., & Schneikert, J. (2012). A common role for various human truncated adenomatous polyposis coli isoforms in the control of beta-catenin activity and cell proliferation. *PLoS One, 7*(4), e34479. doi:10.1371/journal.pone.0034479

Chang, D. T., Pai, R. K., Rybicki, L. A., Dimaio, M. A., Limaye, M., Jayachandran, P., Koong, A. C., Kunz, P. A., Fisher, G. A., Ford, J. M., Welton, M., Shelton, A., Ma, L., Arber, D. A., & Pai, R. K. (2012). Clinicopathologic and molecular features of sporadic early-onset colorectal adenocarcinoma: an adenocarcinoma with frequent signet ring cell differentiation, rectal and sigmoid involvement, and adverse morphologic features. *Mod Pathol, 25*(8), 1128-1139. doi:10.1038/modpathol.2012.61

Charlton, A., Blair, V., Shaw, D., Parry, S., Guilford, P., & Martin, I. G. (2004). Hereditary diffuse gastric cancer: predominance of multiple foci of signet ring cell carcinoma in distal stomach and transitional zone. *Gut, 53*(6), 814-820. doi:10.1136/gut.2002.010447

Cheng, Y. W., Pincas, H., Bacolod, M. D., Schemmann, G., Giardina, S. F., Huang, J., Barral, S., Idrees, K., Khan, S. A., Zeng, Z., Rosenberg, S., Notterman, D. A., Ott, J., Paty, P., & Barany, F. (2008). CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. *Clin Cancer Res, 14*(19), 6005-6013. doi:10.1158/1078-0432.CCR-08-0216

Cheong, C., Oh, S. Y., Kim, Y. B., & Suh, K. W. (2019). Differences in biological behaviors between young and elderly patients with colorectal cancer. *PLoS One, 14*(6), e0218604. doi:10.1371/journal.pone.0218604

Chew, M. H., Koh, P. K., Ng, K. H., & Eu, K. W. (2009). Improved survival in an Asian cohort of young colorectal cancer patients: an analysis of 523 patients from a single institution. *Int J Colorectal Dis, 24*(9), 1075-1083. doi:10.1007/s00384-009-0701-7

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One, 7*(10), e46688. doi:10.1371/journal.pone.0046688

Chou, C. L., Chang, S. C., Lin, T. C., Chen, W. S., Jiang, J. K., Wang, H. S., Yang, S. H., Liang, W. Y., & Lin, J. K. (2011). Differences in clinicopathological characteristics of colorectal cancer between younger and elderly patients: an analysis of 322 patients from a single institution. *Am J Surg, 202*(5), 574-582. doi:10.1016/j.amjsurg.2010.10.014

Christie, M., Jorissen, R. N., Mouradov, D., Sakthianandeswaren, A., Li, S., Day, F., Tsui, C., Lipton, L., Desai, J., Jones, I. T., McLaughlin, S., Ward, R. L., Hawkins, N. J., Ruszkiewicz, A. R., Moore, J., Burgess, A. W., Busam, D., Zhao, Q., Strausberg, R. L., Simpson, A. J., Tomlinson, I. P., Gibbs, P., & Sieber, O. M. (2013). Different APC genotypes in proximal and distal sporadic colorectal cancers suggest distinct WNT/beta-catenin signalling thresholds for tumourigenesis. *Oncogene, 32*(39), 4675-4682. doi:10.1038/onc.2012.486

Christou, N., Perraud, A., Blondy, S., Jauberteau, M. O., Battu, S., & Mathonnet, M. (2017). E-cadherin: A potential biomarker of colorectal cancer prognosis. *Oncol Lett, 13*(6), 4571-4576. doi:10.3892/ol.2017.6063

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin), 6*(2), 80-92. doi:10.4161/fly.19695

Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. *Cell, 127*(3), 469-480. doi:10.1016/j.cell.2006.10.018

Collins, J. E., Bergeland, M. E., Myers, L. L., & Shoop, D. S. (1989). Exfoliating colitis associated with enterotoxigenic Bacteroides fragilis in a piglet. *J Vet Diagn Invest, 1*(4), 349-351. doi:10.1177/104063878900100413

Compton, C., Fenoglio-Preiser, C. M., Pettigrew, N., & Fielding, L. P. (2000). American Joint Committee on Cancer Prognostic Factors Consensus Conference: Colorectal Working Group. *Cancer, 88*(7), 1739-1757. doi:10.1002/(sici)1097-0142(20000401)88:7<1739::aid-cncr30>3.0.co;2-t

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science, 339*(6121), 819-823. doi:10.1126/science.1231143

Conly, J. M., & Stein, K. (1992). The production of menaquinones (vitamin K2) by intestinal bacteria and their role in maintaining coagulation homeostasis. *Prog Food Nutr Sci, 16*(4), 307-343. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1492156

Connell, L. C., Mota, J. M., Braghiroli, M. I., & Hoff, P. M. (2017). The Rising Incidence of Younger Patients With Colorectal Cancer: Questions About Screening, Biology, and Treatment. *Curr Treat Options Oncol, 18*(4), 23. doi:10.1007/s11864-017-0463-3

Cougnoux, A., Dalmasso, G., Martinez, R., Buc, E., Delmas, J., Gibold, L., Sauvanet, P., Darcha, C., Dechelotte, P., Bonnet, M., Pezet, D., Wodrich, H., Darfeuille-Michaud, A., & Bonnet, R. (2014). Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut, 63*(12), 1932-1942. doi:10.1136/gutjnl-2013-305257

Cristofaro, M., Contursi, A., D'Amore, S., Martelli, N., Spaziante, A. F., Moschetta, A., & Villani, G. (2015). Adenomatous polyposis coli (APC)-induced apoptosis of HT29 colorectal cancer cells depends on mitochondrial oxidative metabolism. *Biochim Biophys Acta, 1852*(9), 1719-1728. doi:10.1016/j.bbadis.2015.05.009

Cuevas-Ramos, G., Petit, C. R., Marcq, I., Boury, M., Oswald, E., & Nougayrede, J. P. (2010). Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A, 107*(25), 11537-11542. doi:10.1073/pnas.1001261107

Cunningham, D., Atkin, W., Lenz, H. J., Lynch, H. T., Minsky, B., Nordlinger, B., & Starling, N. (2010). Colorectal cancer. *Lancet, 375*(9719), 1030-1047. doi:10.1016/S0140-6736(10)60353-4

da Silva Tatley, F., Aldwell, F. E., Dunbier, A. K., & Guilford, P. J. (2003). N-terminal E-cadherin peptides act as decoy receptors for Listeria monocytogenes. *Infect Immun, 71*(3), 1580-1583. doi:10.1128/iai.71.3.1580-1583.2003

Darwanto, A., Kitazawa, R., Maeda, S., & Kitazawa, S. (2003). MeCP2 and promoter methylation cooperatively regulate E-cadherin gene expression in colorectal carcinoma. *Cancer Sci, 94*(5), 442-447. doi:10.1111/j.1349-7006.2003.tb01462.x

Daulagala, A. C., Bridges, M. C., & Kourtidis, A. (2019). E-cadherin Beyond Structure: A Signaling Hub in Colon Homeostasis and Disease. *Int J Mol Sci, 20*(11). doi:10.3390/ijms20112756

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature, 505*(7484), 559-563. doi:10.1038/nature12820

Davie, J. R. (2003). Inhibition of histone deacetylase activity by butyrate. *J Nutr, 133*(7 Suppl), 2485S-2493S. doi:10.1093/jn/133.7.2485S

Davies, M. L., Roberts, G. T., Stuart, N., & Wakeman, J. A. (2007). Analysis of a panel of antibodies to APC reveals consistent activity towards an unidentified protein. *Br J Cancer, 97*(3), 384-390. doi:10.1038/sj.bjc.6603873

Davis, D. M., Marcet, J. E., Frattini, J. C., Prather, A. D., Mateka, J. J., & Nfonsam, V. N. (2011). Is it time to lower the recommended screening age for colorectal cancer? *J Am Coll Surg, 213*(3), 352-361. doi:10.1016/j.jamcollsurg.2011.04.033

De Palma, F. D. E., D'Argenio, V., Pol, J., Kroemer, G., Maiuri, M. C., & Salvatore, F. (2019). The Molecular Hallmarks of the Serrated Pathway in Colorectal Cancer. *Cancers (Basel), 11*(7). doi:10.3390/cancers11071017

Deng, G., Bell, I., Crawley, S., Gum, J., Terdiman, J. P., Allen, B. A., Truta, B., Sleisenger, M. H., & Kim, Y. S. (2004). BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin Cancer Res, 10*(1 Pt 1), 191-195. doi:10.1158/1078-0432.ccr-1118-3

Deng, Y. (2017). Rectal Cancer in Asian vs. Western Countries: Why the Variation in Incidence? *Curr Treat Options Oncol, 18*(10), 64. doi:10.1007/s11864-017-0500-2

Derry, M. M., Raina, K., Agarwal, C., & Agarwal, R. (2013). Identifying molecular targets of lifestyle modifications in colon cancer prevention. *Front Oncol, 3*, 119. doi:10.3389/fonc.2013.00119

Desai, T. K., & Barkel, D. (2008). Syndromic colon cancer: lynch syndrome and familial adenomatous polyposis. *Gastroenterol Clin North Am, 37*(1), 47-72, vi. doi:10.1016/j.gtc.2007.12.006

Diether, N. E., & Willing, B. P. (2019). Microbial Fermentation of Dietary Protein: An Important Factor in Diet(-)Microbe(-)Host Interaction. *Microorganisms, 7*(1). doi:10.3390/microorganisms7010019

Dogterom, M., & Koenderink, G. H. (2019). Actin-microtubule crosstalk in cell biology. *Nat Rev Mol Cell Biol, 20*(1), 38-54. doi:10.1038/s41580-018-0067-1

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A, 107*(26), 11971-11975. doi:10.1073/pnas.1002601107

Dozois, E. J., Boardman, L. A., Suwanthanma, W., Limburg, P. J., Cima, R. R., Bakken, J. L., Vierkant, R. A., Aakre, J. A., & Larson, D. W. (2008). Young-onset colorectal cancer in patients with no known genetic predisposition: can we increase early recognition and improve outcome? *Medicine (Baltimore), 87*(5), 259-263. doi:10.1097/MD.0b013e3181881354

Durno, C., Aronson, M., Bapat, B., Cohen, Z., & Gallinger, S. (2005). Family history and molecular features of children, adolescents, and young adults with colorectal carcinoma. *Gut, 54*(8), 1146-1150. doi:10.1136/gut.2005.066092

El-Bahrawy, M., Poulsom, R., Rowan, A. J., Tomlinson, I. T., & Alison, M. R. (2004). Characterization of the E-cadherin/catenin complex in colorectal carcinoma cell lines. *Int J Exp Pathol, 85*(2), 65-74. doi:10.1111/j.0959-9673.2004.0371.x

Esteller, M., Sparks, A., Toyota, M., Sanchez-Cespedes, M., Capella, G., Peinado, M. A., Gonzalez, S., Tarafa, G., Sidransky, D., Meltzer, S. J., Baylin, S. B., & Herman, J. G. (2000). Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer. *Cancer Res, 60*(16), 4366-4371. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10969779

Farrington, S. M., Lin-Goerke, J., Ling, J., Wang, Y., Burczak, J. D., Robbins, D. J., & Dunlop, M. G. (1998). Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. *Am J Hum Genet, 63*(3), 749-759. doi:10.1086/301996

Fearon, E. R. (2011). Molecular genetics of colorectal cancer. *Annu Rev Pathol, 6*, 479-507. doi:10.1146/annurev-pathol-011110-130235

Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell, 61*(5), 759-767. doi:10.1016/0092-8674(90)90186-i

Fedirko, V., Tramacere, I., Bagnardi, V., Rota, M., Scotti, L., Islami, F., Negri, E., Straif, K., Romieu, I., La Vecchia, C., Boffetta, P., & Jenab, M. (2011). Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann Oncol, 22*(9), 1958-1972. doi:10.1093/annonc/mdq653

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., & Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer, 136*(5), E359-386. doi:10.1002/ijc.29210

Fleming, M., Ravula, S., Tatishchev, S. F., & Wang, H. L. (2012). Colorectal carcinoma: Pathologic aspects. *J Gastrointest Oncol, 3*(3), 153-173. doi:10.3978/j.issn.2078-6891.2012.030

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., & Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes, 3*(4), 289-306. doi:10.4161/gmic.19897

Flood, D. M., Weiss, N. S., Cook, L. S., Emerson, J. C., Schwartz, S. M., & Potter, J. D. (2000). Colorectal cancer incidence in Asian migrants to the United States and their descendants. *Cancer Causes Control, 11*(5), 403-411. doi:10.1023/a:1008955722425

Foda, A. A. M., Aziz, A. A., & Mohamed, M. A. (2018). Colorectal signet ring cell carcinoma: Influence of EGFR, E-cadherin and MMP-13 expression on clinicopathological features and prognosis. *Ann Diagn Pathol, 32*, 41-46. doi:10.1016/j.anndiagpath.2017.10.003

Fogh, J. T., G. (1975). New Human Tumor Cell Lines. In J. Fogh (Ed.), *Human Tumor Cells in Vitro* (pp. 115-141). New York: Plenum Publishing Corp.

Fostira, F., & Yannoukakos, D. (2010). A distinct mutation on the alternative splice site of APC exon 9 results in attenuated familial adenomatous polyposis phenotype. *Fam Cancer, 9*(3), 395-400. doi:10.1007/s10689-009-9317-x

Franco, A. A., Mundy, L. M., Trucksis, M., Wu, S., Kaper, J. B., & Sears, C. L. (1997). Cloning and characterization of the Bacteroides fragilis metalloprotease toxin gene. *Infect Immun, 65*(3), 1007-1013. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9038310

Frayling, I. M., Beck, N. E., Ilyas, M., Dove-Edwin, I., Goodman, P., Pack, K., Bell, J. A., Williams, C. B., Hodgson, S. V., Thomas, H. J., Talbot, I. C., Bodmer, W. F., & Tomlinson, I. P. (1998). The APC variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history. *Proc Natl Acad Sci U S A, 95*(18), 10722-10727. doi:10.1073/pnas.95.18.10722

Friedl, W., & Aretz, S. (2005). Familial adenomatous polyposis: experience from a study of 1164 unrelated german polyposis patients. *Hered Cancer Clin Pract, 3*(3), 95-114. doi:10.1186/1897-4287-3-3-95

Fung, K. Y., Lewanowitsch, T., Henderson, S. T., Priebe, I., Hoffmann, P., McColl, S. R., Lockett, T., Head, R., & Cosgrove, L. J. (2009). Proteomic analysis of butyrate effects and loss of butyrate sensitivity in HT29 colorectal cancer cells. *J Proteome Res, 8*(3), 1220-1227. doi:10.1021/pr8009929

Gandhi, J., Davidson, C., Hall, C., Pearson, J., Eglinton, T., Wakeman, C., & Frizelle, F. (2017). Population-based study demonstrating an increase in colorectal cancer in young patients. *Br J Surg, 104*(8), 1063-1068. doi:10.1002/bjs.10518

Gao, M., Zhang, X., Li, D., He, P., Tian, W., & Zeng, B. (2016). Expression analysis and clinical significance of eIF4E, VEGF-C, E-cadherin and MMP-2 in colorectal adenocarcinoma. *Oncotarget, 7*(51), 85502-85514. doi:10.18632/oncotarget.13453

Gausman, V., Dornblaser, D., Anand, S., Hayes, R. B., O'Connell, K., Du, M., & Liang, P. S. (2019). Risk Factors Associated With Early-Onset Colorectal Cancer. *Clin Gastroenterol Hepatol*. doi:10.1016/j.cgh.2019.10.009

Gay, L. J., Mitrou, P. N., Keen, J., Bowman, R., Naguib, A., Cooke, J., Kuhnle, G. G., Burns, P. A., Luben, R., Lentjes, M., Khaw, K. T., Ball, R. Y., Ibrahim, A. E., & Arends, M. J. (2012). Dietary, lifestyle and clinicopathological factors associated with APC mutations and promoter methylation in colorectal cancers from the EPIC-Norfolk study. *J Pathol, 228*(3), 405-415. doi:10.1002/path.4085

Geis, A. L., Fan, H., Wu, X., Wu, S., Huso, D. L., Wolfe, J. L., Sears, C. L., Pardoll, D. M., & Housseau, F. (2015). Regulatory T-cell Response to Enterotoxigenic Bacteroides fragilis Colonization Triggers IL17-Dependent Colon Carcinogenesis. *Cancer Discov, 5*(10), 1098-1109. doi:10.1158/2159-8290.CD-15-0447

Goel, A., Nagasaka, T., Arnold, C. N., Inoue, T., Hamilton, C., Niedzwiecki, D., Compton, C., Mayer, R. J., Goldberg, R., Bertagnolli, M. M., & Boland, C. R. (2007). The CpG island methylator phenotype and chromosomal instability are inversely correlated in sporadic colorectal cancer. *Gastroenterology, 132*(1), 127-138. doi:10.1053/j.gastro.2006.09.018

Gomez-Fernandez, N., Castellvi-Bel, S., Fernandez-Rozadilla, C., Balaguer, F., Munoz, J., Madrigal, I., Mila, M., Grana, B., Vega, A., Castells, A., Carracedo, A., & Ruiz-Ponte, C. (2009). Molecular analysis of the APC and MUTYH genes in Galician and Catalonian FAP families: a different spectrum of mutations? *BMC Med Genet, 10*, 57. doi:10.1186/1471-2350-10-57

Goodwin, A. C., Destefano Shields, C. E., Wu, S., Huso, D. L., Wu, X., Murray-Stewart, T. R., Hacker-Prietz, A., Rabizadeh, S., Woster, P. M., Sears, C. L., & Casero, R. A., Jr. (2011). Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis. *Proc Natl Acad Sci U S A, 108*(37), 15354-15359. doi:10.1073/pnas.1010203108

Gopalan, V., Smith, R. A., Ho, Y. H., & Lam, A. K. (2011). Signet-ring cell carcinoma of colorectum--current perspectives and molecular biology. *Int J Colorectal Dis, 26*(2), 127-133. doi:10.1007/s00384-010-1037-z

Grady, W. M., Willis, J., Guilford, P. J., Dunbier, A. K., Toro, T. T., Lynch, H., Wiesner, G., Ferguson, K., Eng, C., Park, J. G., Kim, S. J., & Markowitz, S. (2000). Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat Genet, 26*(1), 16-17. doi:10.1038/79120

Green, M. R., & Sambrook, J. (2018). The Hanahan Method for Preparation and Transformation of Competent Escherichia coli: High-Efficiency Transformation. *Cold Spring Harb Protoc, 2018*(3). doi:10.1101/pdb.prot101188

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M., & et al. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. *Cell, 66*(3), 589-600. doi:10.1016/0092-8674(81)90021-0

Guilford, P., Blair, V., More, H., & Humar, B. (2007). A short guide to hereditary diffuse gastric cancer. *Hered Cancer Clin Pract, 5*(4), 183-194. doi:10.1186/1897-4287-5-4-183

Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., Taite, H., Scoular, R., Miller, A., & Reeve, A. E. (1998). E-cadherin germline mutations in familial gastric cancer. *Nature, 392*(6674), 402-405. doi:10.1038/32918

Gumbiner, B., Stevenson, B., & Grimaldi, A. (1988). The role of the cell adhesion molecule uvomorulin in the formation and maintenance of the epithelial junctional complex. *J Cell Biol, 107*(4), 1575-1587. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3049625

Gupta, S., Bhattacharya, D., Acharya, A. N., Majumdar, S., Ranjan, P., & Das, S. (2010). Colorectal carcinoma in young adults: a retrospective study on Indian patients: 2000-2008. *Colorectal Dis, 12*(10 Online), e182-189. doi:10.1111/j.1463-1318.2010.02223.x

Haas, S. L., Ye, W., & Lohr, J. M. (2012). Alcohol consumption and digestive tract cancer. *Curr Opin Clin Nutr Metab Care, 15*(5), 457-467. doi:10.1097/MCO.0b013e3283566699

Hakkaart, C., Ellison-Loschmann, L., Day, R., Sporle, A., Koea, J., Harawira, P., Cheng, S., Gray, M., Whaanga, T., Pearce, N., & Guilford, P. (2019). Germline CDH1 mutations are a significant contributor to the high frequency of early-onset diffuse gastric cancer cases in New Zealand Maori. *Fam Cancer, 18*(1), 83-90. doi:10.1007/s10689-018-0080-8

Hall, G., Clarkson, A., Shi, A., Langford, E., Leung, H., Eckstein, R. P., & Gill, A. J. (2010). Immunohistochemistry for PMS2 and MSH6 alone can replace a four antibody panel for mismatch repair deficiency screening in colorectal adenocarcinoma. *Pathology, 42*(5), 409-413. doi:10.3109/00313025.2010.493871

Hamon, M., Bierne, H., & Cossart, P. (2006). Listeria monocytogenes: a multifaceted model. *Nat Rev Microbiol, 4*(6), 423-434. doi:10.1038/nrmicro1413

Han, Y. W., Ikegami, A., Rajanna, C., Kawsar, H. I., Zhou, Y., Li, M., Sojar, H. T., Genco, R. J., Kuramitsu, H. K., & Deng, C. X. (2005). Identification and characterization of a novel adhesin unique to oral fusobacteria. *J Bacteriol, 187*(15), 5330-5340. doi:10.1128/JB.187.15.5330-5340.2005

He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B., & Kinzler, K. W. (1998). Identification of c-MYC as a target of the APC pathway. *Science, 281*(5382), 1509-1512. doi:10.1126/science.281.5382.1509

He, X., Semenov, M., Tamai, K., & Zeng, X. (2004). LDL receptor-related proteins 5 and 6 in Wnt/beta-catenin signaling: arrows point the way. *Development, 131*(8), 1663-1677. doi:10.1242/dev.01117

Hermsen, M., Postma, C., Baak, J., Weiss, M., Rapallo, A., Sciutto, A., Roemen, G., Arends, J. W., Williams, R., Giaretti, W., De Goeij, A., & Meijer, G. (2002). Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology, 123*(4), 1109-1119. doi:10.1053/gast.2002.36051

Hessami Arani, S., & Kerachian, M. A. (2017). Rising rates of colorectal cancer among younger Iranians: is diet to blame? *Curr Oncol, 24*(2), e131-e137. doi:10.3747/co.23.3226

Hill, D. A., Furman, W. L., Billups, C. A., Riedley, S. E., Cain, A. M., Rao, B. N., Pratt, C. B., & Spunt, S. L. (2007). Colorectal carcinoma in childhood and adolescence: a clinicopathologic review. *J Clin Oncol, 25*(36), 5808-5814. doi:10.1200/JCO.2007.12.6102

Hirohashi, S. (1998). Inactivation of the E-cadherin-mediated cell adhesion system in human cancers. *Am J Pathol, 153*(2), 333-339. doi:10.1016/S0002-9440(10)65575-7

Hiscox, S., & Jiang, W. G. (1997). Expression of E-cadherin, alpha, beta and gamma-catenin in human colorectal cancer. *Anticancer Res, 17*(2B), 1349-1354. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9137497

Huang, B., Ni, M., Chen, C., Feng, Y., & Cai, S. (2016). Younger Age Is Associated with Poorer Survival in Patients with Signet-Ring Cell Carcinoma of the Colon without Distant Metastasis. *Gastroenterol Res Pract, 2016*, 2913493. doi:10.1155/2016/2913493

Huet, C., Sahuquillo-Merino, C., Coudrier, E., & Louvard, D. (1987). Absorptive and mucus-secreting subclones isolated from a multipotent intestinal cell line (HT-29) provide new models for cell polarity and terminal differentiation. *J Cell Biol, 105*(1), 345-357. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3611191

Hughes, R., Magee, E. A., & Bingham, S. (2000). Protein degradation in the large intestine: relevance to colorectal cancer. *Curr Issues Intest Microbiol, 1*(2), 51-58. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11709869

Humar, B., Blair, V., Charlton, A., More, H., Martin, I., & Guilford, P. (2009). E-cadherin deficiency initiates gastric signet-ring cell carcinoma in mice and man. *Cancer Res, 69*(5), 2050-2056. doi:10.1158/0008-5472.CAN-08-2457

Hwang, S., Gwon, S. Y., Kim, M. S., Lee, S., & Rhee, K. J. (2013). Bacteroides fragilis Toxin Induces IL-8 Secretion in HT29/C1 Cells through Disruption of E-cadherin Junctions. *Immune Netw, 13*(5), 213-217. doi:10.4110/in.2013.13.5.213

Hyngstrom, J. R., Hu, C. Y., Xing, Y., You, Y. N., Feig, B. W., Skibber, J. M., Rodriguez-Bigas, M. A., Cormier, J. N., & Chang, G. J. (2012). Clinicopathology and outcomes for mucinous and signet ring colorectal adenocarcinoma: analysis from the National Cancer Data Base. *Ann Surg Oncol, 19*(9), 2814-2821. doi:10.1245/s10434-012-2321-7

Inamura, K., Yamauchi, M., Nishihara, R., Kim, S. A., Mima, K., Sukawa, Y., Li, T., Yasunari, M., Zhang, X., Wu, K., Meyerhardt, J. A., Fuchs, C. S., Harris, C. C., Qian, Z. R., & Ogino, S. (2015). Prognostic significance and molecular features of signet-ring cell and mucinous components in colorectal carcinoma. *Ann Surg Oncol, 22*(4), 1226-1235. doi:10.1245/s10434-014-4159-7

Ishida, H., Yamaguchi, T., Chiba, K., Iijima, T., & Horiguchi, S. I. (2017). A case report of ascending colon adenosquamous carcinoma with BRAF V600E mutation. *Int Cancer Conf J, 6*(3), 93-97. doi:10.1007/s13691-017-0283-0

Jenne, D. E., Reimann, H., Nezu, J., Friedel, W., Loff, S., Jeschke, R., Muller, O., Back, W., & Zimmer, M. (1998). Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet, 18*(1), 38-43. doi:10.1038/ng0198-38

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816-821. doi:10.1126/science.1225829

John, B. J., Abulafi, A. M., Poullis, A., & Mendall, M. A. (2007). Chronic subclinical bowel inflammation may explain increased risk of colorectal cancer in obese people. *Gut, 56*(7), 1034-1035. doi:10.1136/gut.2007.125955

Johnson, I. T., & Lund, E. K. (2007). Review article: nutrition, obesity and colorectal cancer. *Aliment Pharmacol Ther, 26*(2), 161-181. doi:10.1111/j.1365-2036.2007.03371.x

Jung, T. H., Park, J. H., Jeon, W. M., & Han, K. S. (2015). Butyrate modulates bacterial adherence on LS174T human colorectal cells by stimulating mucin secretion and MAPK signaling pathway. *Nutr Res Pract, 9*(4), 343-349. doi:10.4162/nrp.2015.9.4.343

Kakar, S., & Smyrk, T. C. (2005). Signet ring cell carcinoma of the colorectum: correlations between microsatellite instability, clinicopathologic features and survival. *Mod Pathol, 18*(2), 244-249. doi:10.1038/modpathol.3800298

Kang, H., O'Connell, J. B., Maggard, M. A., Sack, J., & Ko, C. Y. (2005). A 10-year outcomes evaluation of mucinous and signet-ring cell carcinoma of the colon and rectum. *Dis Colon Rectum, 48*(6), 1161-1168. doi:10.1007/s10350-004-0932-1

Kaurah, P., MacMillan, A., Boyd, N., Senz, J., De Luca, A., Chun, N., Suriano, G., Zaor, S., Van Manen, L., Gilpin, C., Nikkel, S., Connolly-Wilson, M., Weissman, S., Rubinstein, W. S., Sebold, C., Greenstein, R., Stroop, J., Yim, D., Panzini, B., McKinnon, W., Greenblatt, M., Wirtzfeld, D., Fontaine, D., Coit, D., Yoon, S., Chung, D., Lauwers, G., Pizzuti, A., Vaccaro, C., Redal, M. A., Oliveira, C., Tischkowitz, M., Olschwang, S., Gallinger, S., Lynch, H., Green, J., Ford, J., Pharoah, P., Fernandez, B., & Huntsman, D. (2007). Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. *JAMA, 297*(21), 2360-2372. doi:10.1001/jama.297.21.2360

Keenan, J., Aitchison, A., & Frizelle, F. (2017). Are young people eating their way to bowel cancer? *N Z Med J, 130*(1460), 90-92. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28796776

Keenan, J. I., Aitchison, A., Purcell, R. V., Greenlees, R., Pearson, J. F., & Frizelle, F. A. (2016). Screening for enterotoxigenic Bacteroides fragilis in stool samples. *Anaerobe, 40*, 50-53. doi:10.1016/j.anaerobe.2016.05.004

Kim, H., Jen, J., Vogelstein, B., & Hamilton, S. R. (1994). Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *Am J Pathol, 145*(1), 148-156. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8030745

Kim, H. C., Kim, H. J., & Kim, J. C. (2002). Reduced E-cadherin expression as a cause of distinctive signet-ring cell variant in colorectal carcinoma. *J Korean Med Sci, 17*(1), 23-28. doi:10.3346/jkms.2002.17.1.23

Kim, J. M., Cho, S. J., Oh, Y. K., Jung, H. Y., Kim, Y. J., & Kim, N. (2002). Nuclear factor-kappa B activation pathway in intestinal epithelial cells is a major regulator of chemokine gene expression and neutrophil migration induced by Bacteroides fragilis enterotoxin. *Clin Exp Immunol, 130*(1), 59-66. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12296854

Kim, M. C., Kim, C., Wood, L., Neal, D., Kamm, R. D., & Asada, H. H. (2012). Integrating focal adhesion dynamics, cytoskeleton remodeling, and actin motor activity for predicting cell migration on 3D curved surfaces of the extracellular matrix. *Integr Biol (Camb), 4*(11), 1386-1397. doi:10.1039/c2ib20159c

Kim, S., & Coulombe, P. A. (2010). Emerging role for the cytoskeleton as an organizer and regulator of translation. *Nat Rev Mol Cell Biol, 11*(1), 75-81. doi:10.1038/nrm2818

Kinzler, K. W., & Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell, 87*(2), 159-170. doi:10.1016/s0092-8674(00)81333-1

Kirzin, S., Marisa, L., Guimbaud, R., De Reynies, A., Legrain, M., Laurent-Puig, P., Cordelier, P., Pradere, B., Bonnet, D., Meggetto, F., Portier, G., Brousset, P., & Selves, J. (2014). Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. *PLoS One, 9*(8), e103159. doi:10.1371/journal.pone.0103159

Kita, K., Wittmann, T., Nathke, I. S., & Waterman-Storer, C. M. (2006). Adenomatous polyposis coli on microtubule plus ends in cell extensions can promote microtubule net growth with or without EB1. *Mol Biol Cell, 17*(5), 2331-2345. doi:10.1091/mbc.e05-06-0498

Klampfer, L., Huang, J., Sasazuki, T., Shirasawa, S., & Augenlicht, L. (2003). Inhibition of interferon gamma signaling by the short chain fatty acid butyrate. *Mol Cancer Res, 1*(11), 855-862. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14517348

Knudsen, A. L., Bisgaard, M. L., & Bulow, S. (2003). Attenuated familial adenomatous polyposis (AFAP). A review of the literature. *Fam Cancer, 2*(1), 43-55. doi:10.1023/a:1023286520725

Knust, E., & Bossinger, O. (2002). Composition and formation of intercellular junctions in epithelial cells. *Science, 298*(5600), 1955-1959. doi:10.1126/science.1072161

Korpela, K. (2018). Diet, Microbiota, and Metabolic Health: Trade-Off Between Saccharolytic and Proteolytic Fermentation. *Annu Rev Food Sci Technol, 9*, 65-84. doi:10.1146/annurev-food-030117-012830

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., Ojesina, A. I., Jung, J., Bass, A. J., Tabernero, J., Baselga, J., Liu, C., Shivdasani, R. A., Ogino, S., Birren, B. W.,

Huttenhower, C., Garrett, W. S., & Meyerson, M. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res, 22*(2), 292-298. doi:10.1101/gr.126573.111

Kostic, A. D., Howitt, M. R., & Garrett, W. S. (2013). Exploring host-microbiota interactions in animal models and humans. *Genes Dev, 27*(7), 701-718. doi:10.1101/gad.212522.112

Kothari, N., Teer, J. K., Abbott, A. M., Srikumar, T., Zhang, Y., Yoder, S. J., Brohl, A. S., Kim, R. D., Reed, D. R., & Shibata, D. (2016). Increased incidence of FBXW7 and POLE proofreading domain mutations in young adult colorectal cancers. *Cancer, 122*(18), 2828-2835. doi:10.1002/cncr.30082

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc, 4*(7), 1073-1081. doi:10.1038/nprot.2009.86

Lagarde, A., Rouleau, E., Ferrari, A., Noguchi, T., Qiu, J., Briaux, A., Bourdon, V., Remy, V., Gaildrat, P., Adelaide, J., Birnbaum, D., Lidereau, R., Sobol, H., & Olschwang, S. (2010). Germline APC mutation spectrum derived from 863 genomic variations identified through a 15-year medical genetics service to French patients with FAP. *J Med Genet, 47*(10), 721-722. doi:10.1136/jmg.2010.078964

Lamlum, H., Al Tassan, N., Jaeger, E., Frayling, I., Sieber, O., Reza, F. B., Eckert, M., Rowan, A., Barclay, E., Atkin, W., Williams, C., Gilbert, J., Cheadle, J., Bell, J., Houlston, R., Bodmer, W., Sampson, J., & Tomlinson, I. (2000). Germline APC variants in patients with multiple colorectal adenomas, with evidence for the particular importance of E1317Q. *Hum Mol Genet, 9*(15), 2215-2221. doi:10.1093/oxfordjournals.hmg.a018912

Lamlum, H., Ilyas, M., Rowan, A., Clark, S., Johnson, V., Bell, J., Frayling, I., Efstathiou, J., Pack, K., Payne, S., Roylance, R., Gorman, P., Sheer, D., Neale, K., Phillips, R., Talbot, I., Bodmer, W., & Tomlinson, I. (1999). The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. *Nat Med, 5*(9), 1071-1075. doi:10.1038/12511

Larsson, S. C., & Wolk, A. (2007). Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. *Am J Clin Nutr, 86*(3), 556-565. doi:10.1093/ajcn/86.3.556

Laskar, R. S., Talukdar, F. R., Mondal, R., Kannan, R., & Ghosh, S. K. (2014). High frequency of young age rectal cancer in a tertiary care centre of southern Assam, North East India. *Indian J Med Res, 139*(2), 314-318. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24718409

Lee, S., Hwang, K. S., Lee, H. J., Kim, J. S., & Kang, G. H. (2004). Aberrant CpG island hypermethylation of multiple genes in colorectal neoplasia. *Lab Invest, 84*(7), 884-893. doi:10.1038/labinvest.3700108

Li, F. F., Liu, Z., Yan, P., Shao, X., Deng, X., Sam, C., Chen, Y. G., Xu, Y. P., Wang, X. S., Wang, G. Y., & Liu, S. L. (2015). Identification of a novel mutation associated with familial adenomatous polyposis and colorectal cancer. *Int J Mol Med, 36*(4), 1049-1056. doi:10.3892/ijmm.2015.2303

Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Res, 18*(1), 85-98. doi:10.1038/cr.2007.115

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, Q., Cai, G., Li, D., Wang, Y., Zhuo, C., & Cai, S. (2014). Better long-term survival in young patients with non-metastatic colorectal cancer after surgery, an analysis of 69,835 patients in SEER database. *PLoS One, 9*(4), e93756. doi:10.1371/journal.pone.0093756

Liang, J. T., Huang, K. C., Cheng, A. L., Jeng, Y. M., Wu, M. S., & Wang, S. M. (2003). Clinicopathological and molecular biological features of colorectal cancer in patients less than 40 years of age. *Br J Surg, 90*(2), 205-214. doi:10.1002/bjs.4015

Liang, Z., Yan, D., Li, G., & Cheng, H. (2018). Clinical Analysis of Primary Colorectal Signet-Ring Cell Carcinoma. *Clin Colorectal Cancer, 17*(1), e39-e44. doi:10.1016/j.clcc.2017.06.010

Ligtenberg, M. J., Kuiper, R. P., Chan, T. L., Goossens, M., Hebeda, K. M., Voorendt, M., Lee, T. Y., Bodmer, D., Hoenselaar, E., Hendriks-Cornelissen, S. J., Tsui, W. Y., Kong, C. K., Brunner, H. G., van Kessel, A. G., Yuen, S. T., van Krieken, J. H., Leung, S. Y., & Hoogerbrugge, N. (2009). Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet, 41*(1), 112-117. doi:10.1038/ng.283

Limsui, D., Vierkant, R. A., Tillmans, L. S., Wang, A. H., Weisenberger, D. J., Laird, P. W., Lynch, C. F., Anderson, K. E., French, A. J., Haile, R. W., Harnack, L. J., Potter, J. D., Slager, S. L., Smyrk, T. C., Thibodeau, S. N., Cerhan, J. R., & Limburg, P. J. (2010). Cigarette smoking and colorectal cancer risk by molecularly defined subtypes. *J Natl Cancer Inst, 102*(14), 1012-1022. doi:10.1093/jnci/djq201

Linnebacher, M., Wienck, A., Boeck, I., & Klar, E. (2010). Identification of an MSI-H tumor-specific cytotoxic T cell epitope generated by the (-1) frame of U79260(FTO). *J Biomed Biotechnol, 2010*, 841451. doi:10.1155/2010/841451

Littman, D. R., & Pamer, E. G. (2011). Role of the commensal microbiota in normal and pathogenic host immune responses. *Cell Host Microbe, 10*(4), 311-323. doi:10.1016/j.chom.2011.10.004

Liu, P. H., Wu, K., Ng, K., Zauber, A. G., Nguyen, L. H., Song, M., He, X., Fuchs, C. S., Ogino, S., Willett, W. C., Chan, A. T., Giovannucci, E. L., & Cao, Y. (2019). Association of Obesity With Risk of Early-Onset Colorectal Cancer Among Women. *JAMA Oncol, 5*(1), 37-44. doi:10.1001/jamaoncol.2018.4280

Louis, P., & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol, 19*(1), 29-41. doi:10.1111/1462-2920.13589

Louis, P., Hold, G. L., & Flint, H. J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol, 12*(10), 661-672. doi:10.1038/nrmicro3344

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature, 489*(7415), 220-230. doi:10.1038/nature11550

Luchtenborg, M., Weijenberg, M. P., Roemen, G. M., de Bruine, A. P., van den Brandt, P. A., Lentjes, M. H., Brink, M., van Engeland, M., Goldbohm, R. A., & de Goeij, A. F. (2004). APC mutations in sporadic colorectal carcinomas from The Netherlands Cohort Study. *Carcinogenesis, 25*(7), 1219-1226. doi:10.1093/carcin/bgh117

Lynch, H. T., Lanspa, S., Smyrk, T., Boman, B., Watson, P., & Lynch, J. (1991). Hereditary nonpolyposis colorectal cancer (Lynch syndromes I & II). Genetics, pathology, natural history, and cancer control, Part I. *Cancer Genet Cytogenet, 53*(2), 143-160. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1648437

Macfarlane, G. T., Allison, C., Gibson, S. A., & Cummings, J. H. (1988). Contribution of the microflora to proteolysis in the human large intestine. *J Appl Bacteriol, 64*(1), 37-46. doi:10.1111/j.1365-2672.1988.tb02427.x

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science, 339*(6121), 823-826. doi:10.1126/science.1232033

Marshall, B. J., Armstrong, J. A., McGechie, D. B., & Glancy, R. J. (1985). Attempt to fulfil Koch's postulates for pyloric Campylobacter. *Med J Aust, 142*(8), 436-439. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3982345

Masciari, S., Larsson, N., Senz, J., Boyd, N., Kaurah, P., Kandel, M. J., Harris, L. N., Pinheiro, H. C., Troussard, A., Miron, P., Tung, N., Oliveira, C., Collins, L., Schnitt, S., Garber, J. E., & Huntsman, D. (2007). Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet, 44*(11), 726-731. doi:10.1136/jmg.2007.051268

Mauri, G., Sartore-Bianchi, A., Russo, A. G., Marsoni, S., Bardelli, A., & Siena, S. (2019). Early-onset colorectal cancer in young individuals. *Mol Oncol, 13*(2), 109-131. doi:10.1002/1878-0261.12417

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., & Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell, 122*(1), 107-118. doi:10.1016/j.cell.2005.05.007

McCoy, A. N., Araujo-Perez, F., Azcarate-Peril, A., Yeh, J. J., Sandler, R. S., & Keku, T. O. (2013). Fusobacterium is associated with colorectal adenomas. *PLoS One, 8*(1), e53653. doi:10.1371/journal.pone.0053653

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9), 1297-1303. doi:10.1101/gr.107524.110

Mege, R. M., & Ishiyama, N. (2017). Integration of Cadherin Adhesion and Cytoskeleton at Adherens Junctions. *Cold Spring Harb Perspect Biol, 9*(5). doi:10.1101/cshperspect.a028738

Mehenni, H., Lin-Marq, N., Buchet-Poyau, K., Reymond, A., Collart, M. A., Picard, D., & Antonarakis, S. E. (2005). LKB1 interacts with and phosphorylates PTEN: a functional link between two proteins involved in cancer predisposing syndromes. *Hum Mol Genet, 14*(15), 2209-2219. doi:10.1093/hmg/ddi225

Mendonsa, A. M., Na, T. Y., & Gumbiner, B. M. (2018). E-cadherin in contact inhibition and cancer. *Oncogene, 37*(35), 4769-4780. doi:10.1038/s41388-018-0304-2

Mengaud, J., Lecuit, M., Lebrun, M., Nato, F., Mazie, J. C., & Cossart, P. (1996). Antibodies to the leucine-rich repeat region of internalin block entry of Listeria monocytogenes into cells expressing E-cadherin. *Infect Immun, 64*(12), 5430-5433. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8945603

Merino, V. R., Nakano, V., Liu, C., Song, Y., Finegold, S. M., & Avila-Campos, M. J. (2011). Quantitative detection of enterotoxigenic Bacteroides fragilis subtypes isolated from children with and without diarrhea. *J Clin Microbiol, 49*(1), 416-418. doi:10.1128/JCM.01556-10

Mimori-Kiyosue, Y., Shiina, N., & Tsukita, S. (2000). The dynamic behavior of the APC-binding protein EB1 on the distal ends of microtubules. *Curr Biol, 10*(14), 865-868. doi:10.1016/s0960-9822(00)00600-x

Miyaki, M., Konishi, M., Kikuchi-Yanoshita, R., Enomoto, M., Igari, T., Tanaka, K., Muraoka, M., Takahashi, H., Amada, Y., Fukayama, M., & et al. (1994). Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. *Cancer Res, 54*(11), 3011-3020. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8187091

Miyoshi, Y., Nagase, H., Ando, H., Horii, A., Ichii, S., Nakatsuru, S., Aoki, T., Miki, Y., Mori, T., & Nakamura, Y. (1992). Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. *Hum Mol Genet, 1*(4), 229-233. doi:10.1093/hmg/1.4.229

Moghaddam, A. A., Woodward, M., & Huxley, R. (2007). Obesity and risk of colorectal cancer: a meta-analysis of 31 studies with 70,000 events. *Cancer Epidemiol Biomarkers Prev, 16*(12), 2533-2547. doi:10.1158/1055-9965.EPI-07-0708

Molinaro, V., Pensotti, V., Marabelli, M., Feroce, I., Barile, M., Pozzi, S., Laghi, L., Serrano, D., Bernard, L., Bonanni, B., & Ranzani, G. N. (2014). Complementary molecular approaches reveal heterogeneous CDH1 germline defects in Italian patients with hereditary diffuse gastric cancer (HDGC) syndrome. *Genes Chromosomes Cancer, 53*(5), 432-445. doi:10.1002/gcc.22155

Moncrief, J. S., Obiso, R., Jr., Barroso, L. A., Kling, J. J., Wright, R. L., Van Tassell, R. L., Lyerly, D. M., & Wilkins, T. D. (1995). The enterotoxin of Bacteroides fragilis is a metalloprotease. *Infect Immun, 63*(1), 175-181. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7806355

Moore, W. E., Cato, E. P., & Holdeman, L. V. (1978). Some current concepts in intestinal bacteriology. *Am J Clin Nutr, 31*(10 Suppl), S33-42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/707392

Morin, P. J., Vogelstein, B., & Kinzler, K. W. (1996). Apoptosis and APC in colorectal tumorigenesis. *Proc Natl Acad Sci U S A, 93*(15), 7950-7954. doi:10.1073/pnas.93.15.7950

Moseley, J. B., Bartolini, F., Okada, K., Wen, Y., Gundersen, G. G., & Goode, B. L. (2007). Regulated binding of adenomatous polyposis coli protein to actin. *J Biol Chem, 282*(17), 12661-12668. doi:10.1074/jbc.M610615200

Munemitsu, S., Souza, B., Muller, O., Albert, I., Rubinfeld, B., & Polakis, P. (1994). The APC gene product associates with microtubules in vivo and promotes their assembly in vitro. *Cancer Res, 54*(14), 3676-3681. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8033083

Myers, L. L., Firehammer, B. D., Shoop, D. S., & Border, M. M. (1984). Bacteroides fragilis: a possible cause of acute diarrheal disease in newborn lambs. *Infect Immun, 44*(2), 241-244. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/6538870

Myers, L. L., Shoop, D. S., Stackhouse, L. L., Newman, F. S., Flaherty, R. J., Letson, G. W., & Sack, R. B. (1987). Isolation of enterotoxigenic Bacteroides fragilis from humans with diarrhea. *J Clin Microbiol, 25*(12), 2330-2333. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3429625

Nam, J. Y., Oh, B. Y., Hong, H. K., Bae, J. S., Kim, T. W., Ha, S. Y., Park, D., Lee, W. Y., Kim, H. C., Yun, S. H., Park, Y. A., Joung, J. G., Park, W. Y., & Cho, Y. B. (2018). Molecular Characterization of Colorectal Signet-Ring Cell Carcinoma Using Whole-Exome and RNA Sequencing. *Transl Oncol, 11*(4), 836-844. doi:10.1016/j.tranon.2018.04.007

Narayan, S., & Roy, D. (2003). Role of APC and DNA mismatch repair genes in the development of colorectal cancers. *Mol Cancer, 2*, 41. doi:10.1186/1476-4598-2-41

Nathke, I. (2006). Cytoskeleton out of the cupboard: colon cancer and cytoskeletal changes induced by loss of APC. *Nat Rev Cancer, 6*(12), 967-974. doi:10.1038/nrc2010

Nelson-Rees, W. A., Owens, R. B., Arnstein, P., & Kniazeff, A. J. (1976). Source, alterations, characteristics and use of a new dog cell line (Cf2Th). *In Vitro, 12*(10), 665-669. doi:10.1007/BF02797468

Nielsen, M., Hes, F. J., Nagengast, F. M., Weiss, M. M., Mathus-Vliegen, E. M., Morreau, H., Breuning, M. H., Wijnen, J. T., Tops, C. M., & Vasen, H. F. (2007). Germline mutations in APC and MUTYH are responsible for the majority of families with attenuated familial adenomatous polyposis. *Clin Genet, 71*(5), 427-433. doi:10.1111/j.1399-0004.2007.00766.x

Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., & Hedge, P. (1991). Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science, 253*(5020), 665-669. doi:10.1126/science.1651563

Nitsche, U., Zimmermann, A., Spath, C., Muller, T., Maak, M., Schuster, T., Slotta-Huspenina, J., Kaser, S. A., Michalski, C. W., Janssen, K. P., Friess, H., Rosenberg, R., & Bader, F. G. (2013). Mucinous and signet-ring cell colorectal cancers differ from classical adenocarcinomas in tumor biology and prognosis. *Ann Surg, 258*(5), 775-782; discussion 782-773. doi:10.1097/SLA.0b013e3182a69f7e

Norero, E., Alarcon, M. A., Hakkaart, C., de Mayo, T., Mellado, C., Garrido, M., Aguayo, G., Lagos, M., Torres, J., Calvo, A., Guilford, P., & Corvalan, A. H. (2019). Identification of c.1531C>T Pathogenic Variant in the CDH1 Gene as a Novel Germline Mutation of Hereditary Diffuse Gastric Cancer. *Int J Mol Sci, 20*(20). doi:10.3390/ijms20204980

Nosho, K., Irahara, N., Shima, K., Kure, S., Kirkner, G. J., Schernhammer, E. S., Hazra, A., Hunter, D. J., Quackenbush, J., Spiegelman, D., Giovannucci, E. L., Fuchs, C. S., & Ogino, S. (2008). Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample. *PLoS One, 3*(11), e3698. doi:10.1371/journal.pone.0003698

O'Keefe, S. J., Chung, D., Mahmoud, N., Sepulveda, A. R., Manafe, M., Arch, J., Adada, H., & van der Merwe, T. (2007). Why do African Americans get more colon cancer than Native Africans? *J Nutr, 137*(1 Suppl), 175S-182S. doi:10.1093/jn/137.1.175S

O'Keefe, S. J., Li, J. V., Lahti, L., Ou, J., Carbonero, F., Mohammed, K., Posma, J. M., Kinross, J., Wahl, E., Ruder, E., Vipperla, K., Naidoo, V., Mtshali, L., Tims, S., Puylaert, P. G., DeLany, J., Krasinskas, A., Benefiel, A. C., Kaseb, H. O., Newton, K., Nicholson, J. K., de Vos, W. M.,

Gaskins, H. R., & Zoetendal, E. G. (2015). Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun, 6*, 6342. doi:10.1038/ncomms7342

Odamaki, T., Sugahara, H., Yonezawa, S., Yaeshima, T., Iwatsuki, K., Tanabe, S., Tominaga, T., Togashi, H., Benno, Y., & Xiao, J. Z. (2012). Effect of the oral intake of yogurt containing Bifidobacterium longum BB536 on the cell numbers of enterotoxigenic Bacteroides fragilis in microbiota. *Anaerobe, 18*(1), 14-18. doi:10.1016/j.anaerobe.2011.11.004

Ogino, S., Brahmandam, M., Cantor, M., Namgyal, C., Kawasaki, T., Kirkner, G., Meyerhardt, J. A., Loda, M., & Fuchs, C. S. (2006). Distinct molecular features of colorectal carcinoma with signet ring cell component and colorectal carcinoma with mucinous component. *Mod Pathol, 19*(1), 59-68. doi:10.1038/modpathol.3800482

Oliveira, C., Sousa, S., Pinheiro, H., Karam, R., Bordeira-Carrico, R., Senz, J., Kaurah, P., Carvalho, J., Pereira, R., Gusmao, L., Wen, X., Cipriano, M. A., Yokota, J., Carneiro, F., Huntsman, D., & Seruca, R. (2009). Quantification of epigenetic and genetic 2nd hits in CDH1 during hereditary diffuse gastric cancer syndrome progression. *Gastroenterology, 136*(7), 2137-2148. doi:10.1053/j.gastro.2009.02.065

Pan, S. Y., & DesMeules, M. (2009). Energy intake, physical activity, energy balance, and cancer: epidemiologic evidence. *Methods Mol Biol, 472*, 191-215. doi:10.1007/978-1-60327-492-0_8

Paredes, J., Figueiredo, J., Albergaria, A., Oliveira, P., Carvalho, J., Ribeiro, A. S., Caldeira, J., Costa, A. M., Simoes-Correia, J., Oliveira, M. J., Pinheiro, H., Pinho, S. S., Mateus, R., Reis, C. A., Leite, M., Fernandes, M. S., Schmitt, F., Carneiro, F., Figueiredo, C., Oliveira, C., & Seruca, R. (2012). Epithelial E- and P-cadherins: role and clinical significance in cancer. *Biochim Biophys Acta, 1826*(2), 297-311. doi:10.1016/j.bbcan.2012.05.002

Pearlman, R., Markow, M., Knight, D., Chen, W., Arnold, C. A., Pritchard, C. C., Hampel, H., & Frankel, W. L. (2018). Two-stain immunohistochemical screening for Lynch syndrome in colorectal cancer may fail to detect mismatch repair deficiency. *Mod Pathol, 31*(12), 1891-1900. doi:10.1038/s41379-018-0058-y

Perea, J., Alvaro, E., Rodriguez, Y., Gravalos, C., Sanchez-Tome, E., Rivera, B., Colina, F., Carbonell, P., Gonzalez-Sarmiento, R., Hidalgo, M., & Urioste, M. (2010). Approach to early-onset colorectal cancer: clinicopathological, familial, molecular and immunohistochemical characteristics. *World J Gastroenterol, 16*(29), 3697-3703. doi:10.3748/wjg.v16.i29.3697

Perea, J., Cano, J. M., Rueda, D., Garcia, J. L., Inglada, L., Osorio, I., Arriba, M., Perez, J., Gaspar, M., Fernandez-Miguel, T., Rodriguez, Y., Benitez, J., Gonzalez-Sarmiento, R., & Urioste, M. (2015). Classifying early-onset colorectal cancer according to tumor location: new potential subcategories to explore. *Am J Cancer Res, 5*(7), 2308-2313. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26328262

Perea, J., Rueda, D., Canal, A., Rodriguez, Y., Alvaro, E., Osorio, I., Alegre, C., Rivera, B., Martinez, J., Benitez, J., & Urioste, M. (2014). Age at onset should be a major criterion for subclassification of colorectal cancer. *J Mol Diagn, 16*(1), 116-126. doi:10.1016/j.jmoldx.2013.07.010

Perry, J. K., Lins, R. J., Lobie, P. E., & Mitchell, M. D. (2009). Regulation of invasive growth: similar epigenetic mechanisms underpin tumour progression and implantation in human pregnancy. *Clin Sci (Lond), 118*(7), 451-457. doi:10.1042/CS20090503

Phan, Q. T., Myers, C. L., Fu, Y., Sheppard, D. C., Yeaman, M. R., Welch, W. H., Ibrahim, A. S., Edwards, J. E., Jr., & Filler, S. G. (2007). Als3 is a Candida albicans invasin that binds to cadherins and induces endocytosis by host cells. *PLoS Biol, 5*(3), e64. doi:10.1371/journal.pbio.0050064

Pino, M. S., & Chung, D. C. (2010). The chromosomal instability pathway in colon cancer. *Gastroenterology, 138*(6), 2059-2072. doi:10.1053/j.gastro.2009.12.065

Pinto, D., & Clevers, H. (2005). Wnt, stem cells and cancer in the intestine. *Biol Cell, 97*(3), 185-196. doi:10.1042/BC20040094

Plunkett, M., Murray, M., Frizelle, F., Teague, L., Hinder, V., & Findlay, M. (2014). Colorectal adenocarcinoma cancer in New Zealand in those under 25 years of age (1997-2007). *ANZ J Surg, 84*(5), 371-375. doi:10.1111/ans.12380

Polakis, P. (2007). The many ways of Wnt in cancer. *Curr Opin Genet Dev, 17*(1), 45-51. doi:10.1016/j.gde.2006.12.007

Poles, G. C., Clark, D. E., Mayo, S. W., Beierle, E. A., Goldfarb, M., Gow, K. W., Goldin, A., Doski, J. J., Nuchtern, J. G., Vasudevan, S. A., & Langer, M. (2016). Colorectal carcinoma in pediatric patients: A comparison with adult tumors, treatment and outcomes from the National Cancer Database. *J Pediatr Surg, 51*(7), 1061-1066. doi:10.1016/j.jpedsurg.2015.11.005

Polyak, K., & Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer, 9*(4), 265-273. doi:10.1038/nrc2620

Popat, S., Hubner, R., & Houlston, R. S. (2005). Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol, 23*(3), 609-618. doi:10.1200/JCO.2005.01.086

Potten, C. S., Kellett, M., Rew, D. A., & Roberts, S. A. (1992). Proliferation in human gastrointestinal epithelium using bromodeoxyuridine in vivo: data for different sites, proximity to a tumour, and polyposis coli. *Gut, 33*(4), 524-529. doi:10.1136/gut.33.4.524

Poulsen, M. L., & Bisgaard, M. L. (2008). MUTYH Associated Polyposis (MAP). *Curr Genomics, 9*(6), 420-435. doi:10.2174/138920208785699562

Powell, S. M., Zilz, N., Beazer-Barclay, Y., Bryan, T. M., Hamilton, S. R., Thibodeau, S. N., Vogelstein, B., & Kinzler, K. W. (1992). APC mutations occur early during colorectal tumorigenesis. *Nature, 359*(6392), 235-237. doi:10.1038/359235a0

Purcell, R. V., Pearson, J., Aitchison, A., Dixon, L., Frizelle, F. A., & Keenan, J. I. (2017). Colonization with enterotoxigenic Bacteroides fragilis is associated with early-stage colorectal neoplasia. *PLoS One, 12*(2), e0171602. doi:10.1371/journal.pone.0171602

Rahman, R., Schmaltz, C., Jackson, C. S., Simoes, E. J., Jackson-Thompson, J., & Ibdah, J. A. (2015). Increased risk for colorectal cancer under age 50 in racial and ethnic minorities living in the United States. *Cancer Med, 4*(12), 1863-1870. doi:10.1002/cam4.560

Ramamurthy, D., Pazhani, G. P., Sarkar, A., Nandy, R. K., Rajendran, K., Sur, D., Manna, B., & Ramamurthy, T. (2013). Case-control study on the role of enterotoxigenic Bacteroides fragilis as a cause of diarrhea among children in Kolkata, India. *PLoS One, 8*(4), e60622. doi:10.1371/journal.pone.0060622

Raman, R., Kotapalli, V., Adduri, R., Gowrishankar, S., Bashyam, L., Chaudhary, A., Vamsy, M., Patnaik, S., Srinivasulu, M., Sastry, R., Rao, S., Vasala, A., Kalidindi, N., Pollack, J., Murthy, S., & Bashyam, M. (2014). Evidence for possible non-canonical pathway(s) driven early-onset colorectal cancer in India. *Mol Carcinog, 53 Suppl 1*, E181-186. doi:10.1002/mc.21976

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell, 154*(6), 1380-1389. doi:10.1016/j.cell.2013.08.021

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res, 39*(17), e118. doi:10.1093/nar/gkr407

Rhee, K. J., Wu, S., Wu, X., Huso, D. L., Karim, B., Franco, A. A., Rabizadeh, S., Golub, J. E., Mathews, L. E., Shin, J., Sartor, R. B., Golenbock, D., Hamad, A. R., Gan, C. M., Housseau, F., & Sears, C. L. (2009). Induction of persistent colitis by a human commensal, enterotoxigenic Bacteroides fragilis, in wild-type C57BL/6 mice. *Infect Immun, 77*(4), 1708-1718. doi:10.1128/IAI.00814-08

Richardson, A., Hayes, J., Frampton, C., & Potter, J. (2016). Modifiable lifestyle factors that could reduce the incidence of colorectal cancer in New Zealand. *N Z Med J, 129*(1447), 13-20. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27977648

Riegler, M., Lotz, M., Sears, C., Pothoulakis, C., Castagliuolo, I., Wang, C. C., Sedivy, R., Sogukoglu, T., Cosentini, E., Bischof, G., Feil, W., Teleky, B., Hamilton, G., LaMont, J. T., & Wenzl, E. (1999).

Bacteroides fragilis toxin 2 damages human colonic mucosa in vitro. *Gut, 44*(4), 504-510. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10075957

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nat Biotechnol, 29*(1), 24-26. doi:10.1038/nbt.1754

Rosty, C., Williamson, E. J., Clendenning, M., Walters, R. J., Win, A. K., Jenkins, M. A., Hopper, J. L., Winship, I. M., Southey, M. C., Giles, G. G., English, D. R., & Buchanan, D. D. (2014). Should the grading of colorectal adenocarcinoma include microsatellite instability status? *Hum Pathol, 45*(10), 2077-2084. doi:10.1016/j.humpath.2014.06.020

Rowan, A. J., Lamlum, H., Ilyas, M., Wheeler, J., Straub, J., Papadopoulou, A., Bicknell, D., Bodmer, W. F., & Tomlinson, I. P. (2000). APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits". *Proc Natl Acad Sci U S A, 97*(7), 3352-3357. doi:10.1073/pnas.97.7.3352

Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., & Han, Y. W. (2013). Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe, 14*(2), 195-206. doi:10.1016/j.chom.2013.07.012

Sack, R. B., Myers, L. L., Almeido-Hill, J., Shoop, D. S., Bradbury, W. C., Reid, R., & Santosham, M. (1992). Enterotoxigenic Bacteroides fragilis: epidemiologic studies of its role as a human diarrhoeal pathogen. *J Diarrhoeal Dis Res, 10*(1), 4-9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1619243

Salovaara, R., Loukola, A., Kristo, P., Kaariainen, H., Ahtola, H., Eskelinen, M., Harkonen, N., Julkunen, R., Kangas, E., Ojala, S., Tulikoura, J., Valkamo, E., Jarvinen, H., Mecklin, J. P., Aaltonen, L. A., & de la Chapelle, A. (2000). Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J Clin Oncol, 18*(11), 2193-2200. doi:10.1200/JCO.2000.18.11.2193

Sameer, A. S. (2013). Colorectal cancer: molecular mutations and polymorphisms. *Front Oncol, 3*, 114. doi:10.3389/fonc.2013.00114

Samowitz, W. S., Albertsen, H., Sweeney, C., Herrick, J., Caan, B. J., Anderson, K. E., Wolff, R. K., & Slattery, M. L. (2006). Association of smoking, CpG island methylator phenotype, and V600E BRAF mutations in colon cancer. *J Natl Cancer Inst, 98*(23), 1731-1738. doi:10.1093/jnci/djj468

Sanfilippo, L., Baldwin, T. J., Menozzi, M. G., Borriello, S. P., & Mahida, Y. R. (1998). Heterogeneity in responses by primary adult human colonic epithelial cells to purified enterotoxin of Bacteroides fragilis. *Gut, 43*(5), 651-655. doi:10.1136/gut.43.5.651

Sauer, J., Jang, H., Zimmerly, E. M., Kim, K. C., Liu, Z., Chanson, A., Smith, D. E., Mason, J. B., Friso, S., & Choi, S. W. (2010). Ageing, chronic alcohol consumption and folate are determinants of genomic DNA methylation, p16 promoter methylation and the expression of p16 in the mouse colon. *Br J Nutr, 104*(1), 24-30. doi:10.1017/S0007114510000322

Sears, C. L. (2001). The toxins of Bacteroides fragilis. *Toxicon, 39*(11), 1737-1746. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11595636

Sears, C. L. (2009). Enterotoxigenic Bacteroides fragilis: a rogue among symbiotes. *Clin Microbiol Rev, 22*(2), 349-369, Table of Contents. doi:10.1128/CMR.00053-08

Sears, C. L., Geis, A. L., & Housseau, F. (2014). Bacteroides fragilis subverts mucosal biology: from symbiont to colon carcinogenesis. *J Clin Invest, 124*(10), 4166-4172. doi:10.1172/JCI72334

Sears, C. L., & Pardoll, D. M. (2011). Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. *J Infect Dis, 203*(3), 306-311. doi:10.1093/jinfdis/jiq061

Selvanathan, A., Nixon, C. Y., Zhu, Y., Scietti, L., Forneris, F., Uribe, L. M. M., Lidral, A. C., Jezewski, P. A., Mulliken, J. B., Murray, J. C., Buckley, M. F., Cox, T. C., & Roscioli, T. (2020). CDH1 Mutation Distribution and Type Suggests Genetic Differences between the Etiology of Orofacial Clefting and Gastric Cancer. *Genes (Basel), 11*(4). doi:10.3390/genes11040391

Sheffer, M., Bacolod, M. D., Zuk, O., Giardina, S. F., Pincas, H., Barany, F., Paty, P. B., Gerald, W. L., Notterman, D. A., & Domany, E. (2009). Association of survival and disease progression with

chromosomal instability: a genomic exploration of colorectal cancer. *Proc Natl Acad Sci U S A, 106*(17), 7131-7136. doi:10.1073/pnas.0902232106

Shih, I. M., Wang, T. L., Traverso, G., Romans, K., Hamilton, S. R., Ben-Sasson, S., Kinzler, K. W., & Vogelstein, B. (2001). Top-down morphogenesis of colorectal tumors. *Proc Natl Acad Sci U S A, 98*(5), 2640-2645. doi:10.1073/pnas.051629398

Shiozaki, H., Oka, H., Inoue, M., Tamura, S., & Monden, M. (1996). E-cadherin mediated adhesion system in cancer cells. *Cancer, 77*(8 Suppl), 1605-1613. doi:10.1002/(SICI)1097-0142(19960415)77:8<1605::AID-CNCR28>3.0.CO;2-2

Siegel, R. L., Torre, L. A., Soerjomataram, I., Hayes, R. B., Bray, F., Weber, T. K., & Jemal, A. (2019). Global patterns and trends in colorectal cancer incidence in young adults. *Gut, 68*(12), 2179-2185. doi:10.1136/gutjnl-2019-319511

Silla, I. O., Rueda, D., Rodriguez, Y., Garcia, J. L., de la Cruz Vigo, F., & Perea, J. (2014). Early-onset colorectal cancer: a separate subset of colorectal cancer. *World J Gastroenterol, 20*(46), 17288-17296. doi:10.3748/wjg.v20.i46.17288

Simbolo, M., Mafficini, A., Agostini, M., Pedrazzani, C., Bedin, C., Urso, E. D., Nitti, D., Turri, G., Scardoni, M., Fassan, M., & Scarpa, A. (2015). Next-generation sequencing for genetic testing of familial colorectal cancer syndromes. *Hered Cancer Clin Pract, 13*(1), 18. doi:10.1186/s13053-015-0039-9

Sinicrope, F., Foster, N. R., Sargent, D. J., Thibodeau, S. N., Smyrk, T. C., O'Connell, M. J., & North Central Cancer Treatment, G. (2010). Model-based prediction of defective DNA mismatch repair using clinicopathological variables in sporadic colon cancer patients. *Cancer, 116*(7), 1691-1698. doi:10.1002/cncr.24913

Sinicrope, F. A., Rego, R. L., Halling, K. C., Foster, N., Sargent, D. J., La Plant, B., French, A. J., Laurie, J. A., Goldberg, R. M., Thibodeau, S. N., & Witzig, T. E. (2006). Prognostic impact of microsatellite instability and DNA ploidy in human colon carcinoma patients. *Gastroenterology, 131*(3), 729-737. doi:10.1053/j.gastro.2006.06.005

Smith, E. A., & Macfarlane, G. T. (1998). Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. *FEMS Microbiol Ecol, 25*(4), 355-368. doi:10.1111/j.1574-6941.1998.tb00487.x

Soreide, K., Janssen, E. A., Soiland, H., Korner, H., & Baak, J. P. (2006). Microsatellite instability in colorectal cancer. *Br J Surg, 93*(4), 395-406. doi:10.1002/bjs.5328

South, C. D., Yearsley, M., Martin, E., Arnold, M., Frankel, W., & Hampel, H. (2009). Immunohistochemistry staining for the mismatch repair proteins in the clinical care of patients with colorectal cancer. *Genet Med, 11*(11), 812-817. doi:10.1097/GIM.0b013e3181b99b75

Stanich, P. P., Owens, V. L., Sweetser, S., Khambatta, S., Smyrk, T. C., Richardson, R. L., Goetz, M. P., & Patnaik, M. M. (2011). Colonic polyposis and neoplasia in Cowden syndrome. *Mayo Clin Proc, 86*(6), 489-492. doi:10.4065/mcp.2010.0816

Stempelj, M., Kedinger, M., Augenlicht, L., & Klampfer, L. (2007). Essential role of the JAK/STAT1 signaling pathway in the expression of inducible nitric-oxide synthase in intestinal epithelial cells and its regulation by butyrate. *J Biol Chem, 282*(13), 9797-9804. doi:10.1074/jbc.M609426200

Sung, C. O., Seo, J. W., Kim, K. M., Do, I. G., Kim, S. W., & Park, C. K. (2008). Clinical significance of signet-ring cells in colorectal mucinous adenocarcinoma. *Mod Pathol, 21*(12), 1533-1541. doi:10.1038/modpathol.2008.170

Suraweera, N., Duval, A., Reperant, M., Vaury, C., Furlan, D., Leroy, K., Seruca, R., Iacopetta, B., & Hamelin, R. (2002). Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. *Gastroenterology, 123*(6), 1804-1811. doi:10.1053/gast.2002.37070

Takeichi, M. (1990). Cadherins: a molecular family important in selective cell-cell adhesion. *Annu Rev Biochem, 59*, 237-252. doi:10.1146/annurev.bi.59.070190.001321

Takeichi, M. (1991). Cadherin cell adhesion receptors as a morphogenetic regulator. *Science, 251*(5000), 1451-1455. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2006419

Tan, Y., Fu, J., Li, X., Yang, J., Jiang, M., Ding, K., Xu, J., Li, J., & Yuan, Y. (2015). A minor (<50%) signet-ring cell component associated with poor prognosis in colorectal cancer patients: a 26-year retrospective study in China. *PLoS One, 10*(3), e0121944. doi:10.1371/journal.pone.0121944

Testino, G. (2011). The burden of cancer attributable to alcohol consumption. *Maedica (Buchar), 6*(4), 313-320. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22879847

Thiagalingam, S., Laken, S., Willson, J. K., Markowitz, S. D., Kinzler, K. W., Vogelstein, B., & Lengauer, C. (2001). Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc Natl Acad Sci U S A, 98*(5), 2698-2702. doi:10.1073/pnas.051625398

Thibodeau, S. N., Bren, G., & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science, 260*(5109), 816-819. doi:10.1126/science.8484122

Thiele Orberg, E., Fan, H., Tam, A. J., Dejea, C. M., Destefano Shields, C. E., Wu, S., Chung, L., Finard, B. B., Wu, X., Fathi, P., Ganguly, S., Fu, J., Pardoll, D. M., Sears, C. L., & Housseau, F. (2017). The myeloid immune signature of enterotoxigenic Bacteroides fragilis-induced murine colon tumorigenesis. *Mucosal Immunol, 10*(2), 421-433. doi:10.1038/mi.2016.53

Tjalsma, H., Boleij, A., Marchesi, J. R., & Dutilh, B. E. (2012). A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat Rev Microbiol, 10*(8), 575-582. doi:10.1038/nrmicro2819

Toprak, N. U., Yagci, A., Gulluoglu, B. M., Akin, M. L., Demirkalem, P., Celenk, T., & Soyletir, G. (2006). A possible role of Bacteroides fragilis enterotoxin in the aetiology of colorectal cancer. *Clin Microbiol Infect, 12*(8), 782-786. doi:10.1111/j.1469-0691.2006.01494.x

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., & Issa, J. P. (1999). CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A, 96*(15), 8681-8686. doi:10.1073/pnas.96.15.8681

Toyota, M., & Issa, J. P. (1999). CpG island methylator phenotypes in aging and cancer. *Semin Cancer Biol, 9*(5), 349-357. doi:10.1006/scbi.1999.0135

Triantafillidis, J. K., Nasioulas, G., & Kosmidis, P. A. (2009). Colorectal cancer and inflammatory bowel disease: epidemiology, risk factors, mechanisms of carcinogenesis and prevention strategies. *Anticancer Res, 29*(7), 2727-2737. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19596953

Vaahtomeri, K., & Makela, T. P. (2011). Molecular mechanisms of tumor suppression by LKB1. *FEBS Lett, 585*(7), 944-951. doi:10.1016/j.febslet.2010.12.034

van Overbeek, M., Capurso, D., Carter, M. M., Thompson, M. S., Frias, E., Russ, C., Reece-Hoyes, J. S., Nye, C., Gradia, S., Vidal, B., Zheng, J., Hoffman, G. R., Fuller, C. K., & May, A. P. (2016). DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell, 63*(4), 633-646. doi:10.1016/j.molcel.2016.06.037

Vasen, H. F. (2007). Review article: The Lynch syndrome (hereditary nonpolyposis colorectal cancer). *Aliment Pharmacol Ther, 26 Suppl 2*, 113-126. doi:10.1111/j.1365-2036.2007.03479.x

Viljoen, K. S., Dakshinamurthy, A., Goldberg, P., & Blackburn, J. M. (2015). Quantitative profiling of colorectal cancer-associated bacteria reveals associations between fusobacterium spp., enterotoxigenic Bacteroides fragilis (ETBF) and clinicopathological features of colorectal cancer. *PLoS One, 10*(3), e0119462. doi:10.1371/journal.pone.0119462

Vinson, K. E., George, D. C., Fender, A. W., Bertrand, F. E., & Sigounas, G. (2016). The Notch pathway in colorectal cancer. *Int J Cancer, 138*(8), 1835-1842. doi:10.1002/ijc.29800

Vogelaar, I. P., Figueiredo, J., van Rooij, I. A., Simoes-Correia, J., van der Post, R. S., Melo, S., Seruca, R., Carels, C. E., Ligtenberg, M. J., & Hoogerbrugge, N. (2013). Identification of germline mutations in the cancer predisposing gene CDH1 in patients with orofacial clefts. *Hum Mol Genet, 22*(5), 919-926. doi:10.1093/hmg/dds497

Vuik, F. E., Nieuwenburg, S. A., Bardou, M., Lansdorp-Vogelaar, I., Dinis-Ribeiro, M., Bento, M. J., Zadnik, V., Pellise, M., Esteban, L., Kaminski, M. F., Suchanek, S., Ngo, O., Majek, O., Leja, M.,

Kuipers, E. J., & Spaander, M. C. (2019). Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. *Gut, 68*(10), 1820-1826. doi:10.1136/gutjnl-2018-317592

Wang, R., Ma, X., Li, Y., He, Y., Huang, D., Cai, S., & Peng, J. (2016). The Characteristics and Prognostic Effect of E-Cadherin Expression in Colorectal Signet Ring Cell Carcinoma. *PLoS One, 11*(8), e0160527. doi:10.1371/journal.pone.0160527

Wardwell, L. H., Huttenhower, C., & Garrett, W. S. (2011). Current concepts of the intestinal microbiota and the pathogenesis of infection. *Curr Infect Dis Rep, 13*(1), 28-34. doi:10.1007/s11908-010-0147-7

Wei, Q., Wang, X., Gao, J., Li, J., Li, J., Qi, C., Li, Y., Li, Z., & Shen, L. (2016). Clinicopathologic and Molecular Features of Colorectal Adenocarcinoma with Signet-Ring Cell Component. *PLoS One, 11*(6), e0156659. doi:10.1371/journal.pone.0156659

Weisenberger, D. J., Siegmund, K. D., Campan, M., Young, J., Long, T. I., Faasse, M. A., Kang, G. H., Widschwendter, M., Weener, D., Buchanan, D., Koh, H., Simms, L., Barker, M., Leggett, B., Levine, J., Kim, M., French, A. J., Thibodeau, S. N., Jass, J., Haile, R., & Laird, P. W. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet, 38*(7), 787-793. doi:10.1038/ng1834

Wheeler, J. M., Kim, H. C., Efstathiou, J. A., Ilyas, M., Mortensen, N. J., & Bodmer, W. F. (2001). Hypermethylation of the promoter region of the E-cadherin gene (CDH1) in sporadic and ulcerative colitis associated colorectal cancer. *Gut, 48*(3), 367-371. doi:10.1136/gut.48.3.367

White, S., Bubb, V. J., & Wyllie, A. H. (1996). Germline APC mutation (Gln1317) in a cancer-prone family that does not result in familial adenomatous polyposis. *Genes Chromosomes Cancer, 15*(2), 122-128. doi:10.1002/(SICI)1098-2264(199602)15:2<122::AID-GCC7>3.0.CO;2-5

Wick, E. C., & Sears, C. L. (2010). Bacteroides spp. and diarrhea. *Curr Opin Infect Dis, 23*(5), 470-474. doi:10.1097/QCO.0b013e32833da1eb

Willauer, A. N., Liu, Y., Pereira, A. A. L., Lam, M., Morris, J. S., Raghav, K. P. S., Morris, V. K., Menter, D., Broaddus, R., Meric-Bernstam, F., Hayes-Jordan, A., Huh, W., Overman, M. J., Kopetz, S., & Loree, J. M. (2019). Clinical and molecular characterization of early-onset colorectal cancer. *Cancer, 125*(12), 2002-2010. doi:10.1002/cncr.31994

Wodarz, A., & Nathke, I. (2007). Cell polarity in development and cancer. *Nat Cell Biol, 9*(9), 1016-1024. doi:10.1038/ncb433

Wong, J. J., Hawkins, N. J., & Ward, R. L. (2007). Colorectal cancer: a model for epigenetic tumorigenesis. *Gut, 56*(1), 140-148. doi:10.1136/gut.2005.088799

Wu, S., Morin, P. J., Maouyo, D., & Sears, C. L. (2003). Bacteroides fragilis enterotoxin induces c-Myc expression and cellular proliferation. *Gastroenterology, 124*(2), 392-400. doi:10.1053/gast.2003.50047

Wu, S., Rhee, K. J., Zhang, M., Franco, A., & Sears, C. L. (2007). Bacteroides fragilis toxin stimulates intestinal epithelial cell shedding and gamma-secretase-dependent E-cadherin cleavage. *J Cell Sci, 120*(Pt 11), 1944-1952. doi:10.1242/jcs.03455

Xicola, R. M., Li, S., Rodriguez, N., Reinecke, P., Karam, R., Speare, V., Black, M. H., LaDuca, H., & Llor, X. (2019). Clinical features and cancer risk in families with pathogenic CDH1 variants irrespective of clinical criteria. *J Med Genet, 56*(12), 838-843. doi:10.1136/jmedgenet-2019-105991

Yalcin, A. S. A., W.; Yilmaz, A.M.; Aktan, A.O. (2014). Free radicals, whey proteins and colorectal cancer. *Marmara Medical Journal, 27*(1), 1-6. doi:10.5472/MMJ.2013.03144.0

Yamagishi, H., Kuroda, H., Imai, Y., & Hiraishi, H. (2016). Molecular pathogenesis of sporadic colorectal cancers. *Chin J Cancer, 35*, 4. doi:10.1186/s40880-015-0066-y

Yap, A. S., Stevenson, B. R., Keast, J. R., & Manley, S. W. (1995). Cadherin-mediated adhesion and apical membrane assembly define distinct steps during thyroid epithelial polarization and lumen formation. *Endocrinology, 136*(10), 4672-4680. doi:10.1210/endo.136.10.7664688

Appendix 1 PCR Primer Sequences

Table A1. *CDH1* amplicon specific primers for next generation sequencing

Name	Sequence (5'-3')	Product Size (bp)
CDH1 Prom F	ACGACGCTCTTCCGATCTTCGAACCCAGTGGAATCAGAAC	300
CDH1 Prom R	CGTGTGCTCTTCCGATCTACAGGTGCTTTGCAGTTCCG	
CDH1 Exon 1 F	ACGACGCTCTTCCGATCTGAACTGCAAAGCACCTGTGA	286
CDH1 Exon 1 R	CGTGTGCTCTTCCGATCTGTGACGACGGGAGAGGAAG	
CDH1 Exon 2 F	ACGACGCTCTTCCGATCTTTTCGGTGAGCAGGAGGGAA	291
CDH1 Exon 2 R	CGTGTGCTCTTCCGATCTGGTGTGGGAGTGCAATTTCT	
CDH1 Exon 3 F	ACGACGCTCTTCCGATCTCGCTCTTTGGAGAAGGAATG	410
CDH1 Exon 3 R	CGTGTGCTCTTCCGATCTCGGTACCAAGGCTGAGAAAC	
CDH1 Exon 4 F	ACGACGCTCTTCCGATCTTGATTGGTCATTTTGGTGGA	460
CDH1 Exon 4 R	CGTGTGCTCTTCCGATCTGAATTAGTAAAGAAGGATCCCAAC	
CDH1 Exon 5 F	ACGACGCTCTTCCGATCTAGTGTTGGGATCCTTCTT	393
CDH1 Exon 5 R	CGTGTGCTCTTCCGATCTCCCATCACTTCTCCTTAGCA	
CDH1 Exon 6 F	ACGACGCTCTTCCGATCTCAGCAGCACATGTGTGAGAAAAGTC	440
CDH1 Exon 6 R	CGTGTGCTCTTCCGATCTGGAAGGATCAGCTTTAGTTACAC	
CDH1 Exon 7 F	ACGACGCTCTTCCGATCTCCAGTCCCAAAGTGCAGCTTGTCT	292
CDH1 Exon 7 R	CGTGTGCTCTTCCGATCTCACCCTCTGGATCCTCCTGA	
CDH1 Exon 8 F	ACGACGCTCTTCCGATCTGTTCCGTGCCTAGAAGACA	452
CDH1 Exon 8 R	CGTGTGCTCTTCCGATCTGCCATCTCAAGATGCTTGCT	
CDH1 Exon 9 F	ACGACGCTCTTCCGATCTTGACACATCTCTTTGCTCTGC	318
CDH1 Exon 9 R	CGTGTGCTCTTCCGATCTAGAAGATACCAGGGGACAAGG	
CDH1 Exon 10 F	ACGACGCTCTTCCGATCTAACCACAGTTACTTTTGCACC	445
CDH1 Exon 10 R	CGTGTGCTCTTCCGATCTAACCAGTTGCTGCAAGTCAG	
CDH1 Exon 11 F	ACGACGCTCTTCCGATCTTTCTAAAAGCCAGAGCTTGTCC	277
CDH1 Exon 11 R	CGTGTGCTCTTCCGATCTGAGGGGCAAGGAACTGAACT	
CDH1 Exon 12 F	ACGACGCTCTTCCGATCTACCACTGAAGAGCCAGGAC	444
CDH1 Exon 12 R	CGTGTGCTCTTCCGATCTGAAATTGAAAGGTGGGGATCT	
CDH1 Exon 13 F	ACGACGCTCTTCCGATCTCGGGTGTCTTTAGTTCACTAGC	484
CDH1 Exon 13 R	CGTGTGCTCTTCCGATCTTGGGAGTCTCTTTCCCACAT	
CDH1 Exon 14 F	ACGACGCTCTTCCGATCTGTGATAGCTGCTGCTTCTGG	324
CDH1 Exon 14 R	CGTGTGCTCTTCCGATCTTGTTTCAAATGCCTACCTC	
CDH1 Exon 15 F	ACGACGCTCTTCCGATCTACATAGCCCTGTGTGTATGAC	289
CDH1 Exon 15 R	CGTGTGCTCTTCCGATCTAGAGATGAGCCATGCTTTGG	
CDH1 Exon 16 F	ACGACGCTCTTCCGATCTGATGACAGGTGTGCCCTTC	363
CDH1 Exon 16 R	CGTGTGCTCTTCCGATCTCAGCAACGTGATTTCTGCAT	

Amplicon-specific primers (black bases) designed with non-specific 18bp sequences at the 5' end of primers (red bases). Two non-specific sequences used, one for forward primers, one for reverse primers. bp, base pairs.

Table A2. *APC* amplicon specific primers for next generation sequencing

Name	Sequence (5'-3')	Product Size (bp)
APC Exon 5 F	ACGACGCTCTTCCGATCTCACCATGACTGACGTATTTGCT	389
APC Exon 5 R	CGTGTGCTCTTCCGATCTAGAGCCAAAATAAACACAGCCTT	
APC Exon 6 F	ACGACGCTCTTCCGATCTCCTGAGCTTTTAAGTGGTAGCC	458
APC Exon 6 R	CGTGTGCTCTTCCGATCTTGTAAACTGACAGCTAAAGTAAGGT	
APC Exon 8 F	ACGACGCTCTTCCGATCTTCTGCAGTTTAATGCTCATATGCAA	402
APC Exon 8 R	CGTGTGCTCTTCCGATCTTGGCATTAGTGACCAGGGTTT	
APC Exon 9 F1	ACGACGCTCTTCCGATCTTCATCACTTAATTGGTTTTTGGCTT	397
APC Exon 9 R1	CGTGTGCTCTTCCGATCTAAGGACTCGGATTTCACGCC	
APC Exon 9 F2	ACGACGCTCTTCCGATCTGGAAATTCCCGGGGCAGTAA	374
APC Exon 9 R2	CGTGTGCTCTTCCGATCTTGAGTAGCACAAATGGCTGA	
APC Exon 11 F1	ACGACGCTCTTCCGATCTGGGGTGGAGAAACTGGCATA	406
APC Exon 11 R1	CGTGTGCTCTTCCGATCTACCTTGTTGGCTACATCTCCAAA	
APC Exon 11 F2	ACGACGCTCTTCCGATCTTAGGGGGACTACAGGCCATT	320
APC Exon 11 R2	CGTGTGCTCTTCCGATCTGCGAATGTGAAGCACAGGTT	
APC Exon 12 F	ACGACGCTCTTCCGATCTTGGCTTCAAGTTGTCTTTTTAATG	511
APC Exon 12 R	CGTGTGCTCTTCCGATCTTGAGATACTAAATACTGAGCAACAA	
APC Exon 13 F	ACGACGCTCTTCCGATCTAGTGATAGGATTACAGGCGTG	476
APC Exon 13 R	CGTGTGCTCTTCCGATCTGAAATTAGGGAATCTCATGG	
APC Exon 15 F1	ACGACGCTCTTCCGATCTTCCCAAGGCATCTCATCGT	387
APC Exon 15 R1	CGTGTGCTCTTCCGATCTATGGCTGACACTTCTTCCATGA	
APC Exon 15 F2	ACGACGCTCTTCCGATCTAGAAGCTCTGCTGCCCATAC	387
APC Exon 15 R2	CGTGTGCTCTTCCGATCTCCTTCCAGAGTTCAACTGCTCA	
APC Exon 15 F3	ACGACGCTCTTCCGATCTTCCATACAGGTCACGGGGAG	436
APC Exon 15 R3	CGTGTGCTCTTCCGATCTTGGATGGAGCTGATTCTGCC	
APC Exon 15 F4	ACGACGCTCTTCCGATCTAGCAGTGAGAATACGTCCACA	449
APC Exon 15 R4	CGTGTGCTCTTCCGATCTGCTTTGTGCCTGGCTGATTC	
APC Exon 15 F5	ACGACGCTCTTCCGATCTGCAGACTGCAGGGTTCTAGT	470
APC Exon 15 R5	CGTGTGCTCTTCCGATCTGACCCTCTGAACTGCAGCAT	
APC Exon 15 F6	ACGACGCTCTTCCGATCTAACAGCTCAAACCAAGCGAG	504
APC Exon 15 R6	CGTGTGCTCTTCCGATCTTTTGAAGCAGTCTGGGCTGG	

Amplicon-specific primers (black bases) designed with non-specific 18bp sequences at the 5' end of primers (red bases). Two non-specific sequences used, one for forward primers, one for reverse primers. bp, base pairs.

Table A3. Forward adapter primers for creating next generation sequencing libraries

Name	Sequence (5'-3')
FPI1	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **CGTGAT** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI2	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **ACATCG** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI3	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **GCCTAA** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI4	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **TGGTCA** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI5	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **CACTGT** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI6	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **ATTGGC** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI7	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **GATCTG** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI8	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **TCAAGT** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI9	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **CTGATC** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI10	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **AAGCTA** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI11	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **GTAGCC** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI12	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **TACAAG** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI13	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **TTGACT** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI14	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **GGAACT** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI15	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **TGACAT** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T
FPI16	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC **GGACGG** ACAC T CTT TCC CTA CAC GAC GCT CTT CCG ATC T

Forward adapter primers with unique index sequences shown in **red**.

Table A4. Reverse adapter primers for creating next generation sequencing libraries

Name	Sequence (5'-3')
RPI1	CAA GCA GAA GAC GGC ATA CGA GAT **CGTGAT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI2	CAA GCA GAA GAC GGC ATA CGA GAT **ACATCG** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI3	CAA GCA GAA GAC GGC ATA CGA GAT **GCCTAA** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI4	CAA GCA GAA GAC GGC ATA CGA GAT **TGGTCA** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI5	CAA GCA GAA GAC GGC ATA CGA GAT **CACTGT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI6	CAA GCA GAA GAC GGC ATA CGA GAT **ATTGGC** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI7	CAA GCA GAA GAC GGC ATA CGA GAT **GATCTG** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI8	CAA GCA GAA GAC GGC ATA CGA GAT **TCAAGT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI9	CAA GCA GAA GAC GGC ATA CGA GAT **CTGATC** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI10	CAA GCA GAA GAC GGC ATA CGA GAT **AAGCTA** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI11	CAA GCA GAA GAC GGC ATA CGA GAT **GTAGCC** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI12	CAA GCA GAA GAC GGC ATA CGA GAT **TACAAG** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI13	CAA GCA GAA GAC GGC ATA CGA GAT **TTGACT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI14	CAA GCA GAA GAC GGC ATA CGA GAT **GGAACT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI15	CAA GCA GAA GAC GGC ATA CGA GAT **TGACAT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
RPI16	CAA GCA GAA GAC GGC ATA CGA GAT **GGACGG** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC

Reverse adapter primers with unique index sequences shown in **red**.

Table A5 Primers for generating template for CRISPR/Cas9n editing

Name	Sequence (5'-3')	Restriction site
HT29_CRISPR_Template1F	GCGGTACCTGTTTTTGACACCAATCG	KpnI
HT29_CRISPR_Template1R	GCGCGGCCGCTTGGCAATCTGGGCTGCAG	NotI

Primers to amplify the region of APC surrounding the mutation to be corrected. Primers incorporate KpnI and NotI restriction sites, respectively (blue bases), for digestion to allow ligation into pBluescript KS+ plasmid.

Table A6 Primers for mutating template for CRISPR/Cas9n editing

Name	Sequence (5'-3')	Restriction Site
APC template1 mut1F	GCTCCTCTTCATCTAGAGG	XbaI
APC template1 mut1R	CCTCTAGATGAAGAGGAGC	
APC template1 mut2F	GAAGTCTCGAGAGAGAACG	XhoI
APC template1 mut2R	CGTTCTCTCTCGAGACTTC	

Primers designed to introduce XbaI and XhoI restriction sites, respectively, into the template plasmid (red bases differ from genomic sequence) without altering the protein code.

Table A7 Primers for generating gRNA recognition sequences

Name	Sequence (5'-3')
HT29_APC_gRNAn1AF	CACCGTCCTCTTGATGAAGAGGAGC
HT29_APC_gRNAn1AR	AAACGCTCCTCTTCATCAAGAGGAC
HT29_APC_gRNAn1BF	CACCGAGAAGTTTGTAGAGAGAACG
HT29_APC_gRNAn1BR	AAACCGTTCTCTCTACAAACTTCTC

Pairs of primers with complementary sequences that anneal to form gRNA recognition sequences that recognise *APC* sequences in the region of the mutated base. Green bases denote sequences that recognize BpiI digested DNA, brown bases were added as the Cas9n system requires recognition sequences to start with a guanine nucleotide, red bases denote the mutated base.

Appendix 2 Buffers and Solutions

2 x protein loading buffer (reducing): 0.4 g of sodium dodecyl sulphate (SDS) was dissolved in 7 mL of 0.125 M Tris solution (pH6.8). Two mL of glycerol and 0.002 g of bromophenol blue were added along with 1 mL of β-mercaptoethanol. The buffer was stored in 1 mL aliquots at -20°C.

6 x DNA loading buffer: A 40 % (w/v) solution of sucrose was made with 0.25% (w/v) bromophenol blue and xylene cyanol. 1 mL aliquots were stored at -20°C.

Phosphate Buffered Saline (PBS): A 10 x stock solution was made by dissolving 80 g sodium chloride, 2 g potassium chloride, 28.42 g di-sodium hydrogen phosphate and 2 g potassium dihydrogen phosphate in 1 L MPW. The stock was diluted 1:10 in mqH_2O as required.

PIPES buffer: 9.07 g of 1,4-piperazinediethanesulfonic acid (PIPES) were dissolved in 400 mL MPW. The pH was adjusted to 6.8 with sodium hydroxide and the total volume adjusted to 500 mL.

Polyacrylamide gel loading buffer (top): A 10 x stock solution was made by dissolving 60 g of Tris base, 286 g of glycine and 20 g of SDS in MPW to a total volume of 2 L. A working solution was made by diluting the stock 1:10 in MPW.

Polyacrylamide gel loading buffer (bottom): A 2 L working solution was made up by dissolving 20 g of Tris base in MPW.

Protein lysis buffer: A solution of 100 mM sodium chloride, 5mM EDTA, 1% (v/v) Nonidet P-40, 0.2% (w/v) SDS, 1 mM phenylmethanesulfonyl fluoride (PMSF), 1 mM sodium orthovanadate, and 0.5% (w/v) sodium deoxycholate was made in 50 mM tris-HCl to a total volume of 10 mL. One cOmplete mini protease inhibitor cocktail tablet was dissolved in the buffer which was then stored in 1 mL aliquots at -20°C until use.

TBE (Tris Borate EDTA): A 10 x stock solution was made up by dissolving 108 g of Tris base 55 g of boric acid and 7.5 g of EDTA in MPW to a total volume of 1 L. The stock was diluted 1:10 in MPW as required.

TBST: A 10 x stock solution of Tris buffered saline (TBS) was made by dissolving 15.125 g of Tris base and 44 g of sodium chloride in 450 mL of MPW, the pH adjusted to 7.5 with hydrochloric acid and the volume adjusted to 500 mL. The solution was diluted 1:10 in MPW and 1 mL Tween 20 detergent added to make a working solution.

Transfer buffer: A 10 x stock solution was made by dissolving 60.4 g of Tris base and 288 g of glycine in MPW to a total volume of 2 L. Working solutions were made by adding 100 mL of 10 x stock solution to 800 mL of MPW and 100 mL of methanol. Working solution was stored at 4°C.

Tris-HCl buffer: A 0.5M Tris solution was made up by dissolving 3 g of Tris base in 40 mL of mqH_2O and the pH adjusted to 6.8 with hydrochloric acid. The final volume was adjusted to 50 mL with MPW.

Tris-EDTA buffer: Ten microliters of 0.5 M EDTA solution was added to 5 mL of 0.5 M Tris-HCl buffer (pH6.8) and the final volume adjusted to 50 mL to give final concentrations of 10 mM Tris and 0.1 mM EDTA.

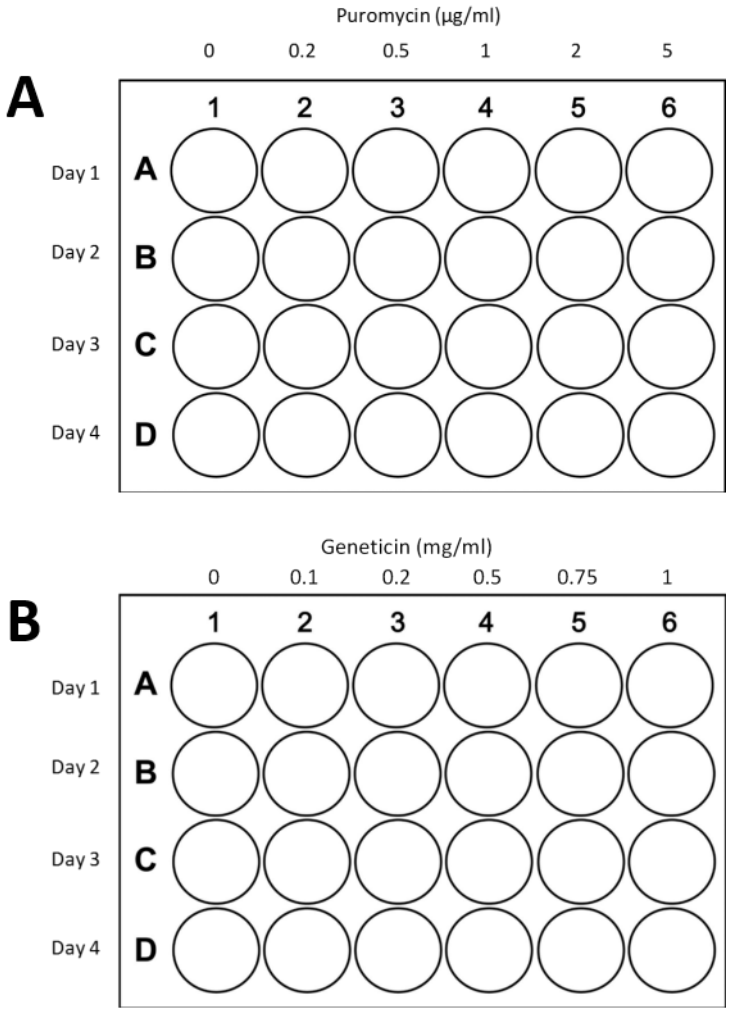

Figure A1. Drug concentrations for selective drug sensitivity testing. To test the sensitivity of HT29 cells to the selective drugs puromycin (A) and geneticin (B), cells were grown in wells of a 24-well culture plate in the presence of drugs at the above concentrations (Section 2.2.4.9). Note the differences in units between puromycin and geneticin.