# The Genetics and Pharmacogenetics of Colorectal Cancer

Submitted for the degree of Doctor of Philosophy
at Cardiff University

**Christopher G. Smith**

**2011**

UMI Number: U584542

UMI

Dissertation Publishing

UMI U584542

ProQuest

## DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ...................................... (candidate)    Date ..21/09/2011........

## STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of ......PhD.....................(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed ...................................... (candidate)    Date ..21/09/2011.....

## STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources are acknowledged by explicit references.

Signed ...................................... (candidate)    Date ..21/09/2011.....

## STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ...................................... (candidate)    Date ..21/09/2011.....

## STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access previously approved by the Graduate Development Committee.**

Signed ...................................... (candidate)    Date ...........................

## Summary

We aimed to identify genetic factors that alter colorectal cancer (CRC) risk, patient survival and response to treatment. Using a cohort of 2,186 advanced CRC (aCRC) cases and 2,176 healthy controls, we validated a role for loci at 18q21, 8q23, 15q13 and 8q24 in CRC susceptibility. We also identified a variant in *OGG1* that may act as a novel low penetrance risk allele. We identified four germline SNPs from 14 genome wide associated CRC risk loci that independently altered patient survival (rs16892766 at 8q23, HR 1.28, 95% CI 1.13-1.45, P<0.001; rs9929218 at 16q22, HR 1.46, 95% CI 1.23-1.74, P=0.002; rs10411210 at 19q13, HR 1.23, 95% CI 1.08-1.39, P=0.001; and, rs10795668 at 10p14, HR 0.71, 95% CI 0.59-0.86, P=0.001). We found that somatic mutations of the *KRAS* (HR 1.34, 95% CI 1.18-1.52 P<0.001) and *BRAF* (HR 2.00, 95% CI 1.61-2.50, P<0.001) oncogenes altered patient survival independent of treatment. Through the COIN trial, we found no evidence for improved response to cetuximab in patients' wild-type for *KRAS* (OS; HR 1.04, 95% CI 0.87-1.23, P=0.67) or all wild-type for *KRAS, BRAF* and *NRAS* (OS; HR 1.02, 95% CI 0.83-1.24, P=0.86). Similarly, *PIK3CA* mutation status did not correlate with response (OS; 1.03, 95% CI 0.86-1.24, P=0.89). However, we identified a 'most responsive' cohort (patients wild-type for *KRAS, BRAF* and *NRAS*, with ≤1 metastatic sites and that received OxFU) that showed improved response (PFS; HR 0.55, 95% CI 0.35-0.87, P=0.01). Analysis of individual somatic variants revealed no significant associations with response. We found that rs9929218 altered response to (OR 0.48, 95% CI 0.31-0.75, P=0.001) and side effects from (OR 4.68, 95% CI 1.84-11.9, P=0.001) standard chemotherapy and response to cetuximab in a *KRAS, BRAF* and *NRAS* mutant dependent manner (OR 1.69, 95% CI 0.61-4.74, P for interaction 0.004). We elucidated the underlying mechanisms at the 8q23 and 16q22 loci. These data may help to tailor future therapies for patients with aCRC.

# Acknowledgements

# Abbreviations

| | |
|---|---|
| 3C | Chromosome Confirmation Capture |
| 5'-DFCR | 5'-deoxy-5-fluorocytidine |
| 5'-DFUR | 5'-deoxy-fluorouridine |
| 5FU | 5-Fluorouracil |
| 8oxoG | 8-oxo-7,8-dihydro2'deoxyguanosine |
| 95% CI | 95% Confidence Interval |
| ABC | Avidin-Biotin Complex |
| aCRC | Advanced Colorectal Cancer |
| ADCC | Antibody-dependent Cell-mediated Cytotoxicity |
| AFAP | Attenuated-Familial Adenomatous Polyposis |
| AP Site | Apurinic/Apyrimidinic Site |
| APC | Adenomatous Polyposis Coli |
| ASO | Anti-Sense Oligonucleotide |
| B2M | Beta 2-Microglobulin |
| BER | Base Excision Repair |
| CIMP | CpG Island Methylator Phenotype |
| CIN | Chromosomal Instability |
| CIP | Calf Intestinal Alkaline Phosphatase |
| CK1 | Casein Kinase 1α |
| CMV | Cytomegalovirus |
| COIN | COntinuous vs. INtermittent |
| COOH | Carboxy |
| COSMIC | Catalogue of Somatic Mutations in Cancer |
| CR | Complete Response |
| CRA | Colorectal Adenoma |
| CRC | Colorectal Cancer |
| CRF | Clinical Return Form |
| CTS | Contents Trade Secret |
| CTU | Clinical Trials Unit |
| DAB | 3,3'-diaminobenzidine |
| ddNTP | dideoxynucleotide triphosphate |
| DGGE | Denature Gradient Gel Electrophoresis |
| DMEM | Dulbecco's modified eagle's medium |
| DNA | Deoxyribonucleic Acid |
| dNTP | Deoxynucleotidetriphosphate |
| DP | Disease Progression |
| DPD | Dihydropyrimidine Dehydrogenase |
| DSBR | Double Strand Break Repair |
| Dsh | Dishevelled |
| EGFR | Epidermal Growth Factor Receptor |
| EIF3H | Eukaryotic Initation Factor 3 H |
| EIF4E | Eukaryotic Initation Factor 4 E |
| EMSA | Electrophorectic Mobility Shift Assay |
| EMT | Epithelial to Mesenchymal Transition |
| EXO1 | Exonuclease 1 |
| FAP | Familial Adenomatous Polyposis |
| FBS | Fetal bovine serum |
| FdUMP | Fluorodeoxyuridine Monophosphate |
| FEN1 | Flap Endonuclease 1 |
| FFPE | Formalin Fixed Paraffin-Embedded |
| FOL | Folinic Acid |

| FP | Fractional Polynomial |
| Frz | Frizzled |
| GFP | Green Fluorescent protein |
| GI | Gastrointestinal |
| GSK3 | Glycosen Synthase Kinase 3 |
| GST | Glutathion S-Transferase |
| GWAS | Genome Wide Association Study/Scan |
| H&E | Haematoxylin and Eosin |
| HFS | Hand Foot Syndrome |
| HNPCC | Hereditary Non-Polyposis Colorectal Cancer |
| HPLC | High Performance Liquid Chromatography |
| HPSF | High purity salt free |
| HR | Hazard Ratio |
| HWE | Hardy Weinberg Equilibrium |
| IHC | Immunohistochemistry |
| IPTG | Isopropyl-$\beta$-D-thio-galactopyranoside |
| IV | Intravenous |
| LAR II | Luciferase Assay Reagent II |
| LCM | Laser Capture Microdissection |
| LD | Linkage Disequilibrium |
| LOH | Loss of Heterozygosity |
| MAF | Minor Allele Frequency |
| MALDI-TOF | Matrix Assisted Laser Desorption Ionisation - Time of Flight |
| MAP | *MUTYH*-Associated Polyposis |
| MCR | Mutation Cluster Region |
| mCRC | Metastatic Colorectal Cancer |
| MMR | Mistmatch Repair |
| mRNA | Messenger RNA |
| MSI | Microsatellite Instability |
| MSS | Microsatellite Stable |
| n | Number |
| NER | Nucleotide Excision Repair |
| NGS | Next Generation Sequencing |
| NMD | Nonsense Mediated Decay |
| nsSNP | nonsynonymous SNP |
| ODC | Ornithine Decarboxylase |
| OR | Odds Ratio |
| ORF | Open Reading Frame |
| OS | Overall Survival |
| OxCap | Oxaliplatin and Capecitabine therapy |
| OxDR | Oxidative Damage Repair |
| OxFp | Oxaliplatin and a Fluoropyrimidine |
| OxFU | Oxaliplatin and 5FU therapy |
| PBS | Phospho-buffered Saline |
| PCR | Polymerase Chain Reaction |
| PFS | Progression Free Survival |
| POL$\delta$ | Polymerase $\delta$ |
| RACE | Rapid Amplification of cDNA Ends |
| RCT | Randomised Controlled Trial |
| RFC | Replication Factor C |
| RHPN2 | Rho GTPase Binding Protein 2 |
| RL | Renilla |

| | |
|---|---|
| RNA | Ribonucleic Acid |
| ROS | Reactive Oxygen Species |
| RSJ | Recto-Sigmoid Junction |
| RT-PCR | Reverse Transcription-PCR |
| SAP | Shrimp Alkaline Phosphatase |
| SD | Stable Disease |
| SMAD7 | Mothers Against Decapentaplegic Homolog 7 |
| SNP | Single Nucleotide Polymorphism |
| SSCP | Single-Strand Conformation Polymorphism |
| TGF | Transforming Growth Factor |
| Topo I | Topoisomerase I |
| TP | Thymidine phosphorylase |
| TP | Training Phase |
| TS | Thymidylate synthase |
| UKBS | UK Blood Service |
| uORF | Upstream Open Reading Frame |
| UV | Ultraviolet |
| VEGF | Vascular Endothelial Growth Factor |
| VP | Validation Phase |
| WBC | White Blood Cell |
| WHO-PS | World Health Organisation Performance Status |
| wt | Wild-Type |
| X-gal | 5-bromo-4-chloro-3-indoyl-D-galactoside |
| XRCC1 | X-ray Repair Cross Complementing 1 |

# Codon Table

| | | Second position of codon | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | T | | C | | A | | G | | | |
| First position | T | TTT | F | TCT | S | TAT | Y | TGT | C | T | Third position |
| | | TTC | F | TCC | S | TAC | Y | TGC | C | C | |
| | | TTA | L | TCA | S | TAA | Stop | TGA | Stop | A | |
| | | TTG | L | TCG | S | TAG | Stop | TGG | W | G | |
| | C | CTT | L | CCT | P | CAT | H | CGT | R | T | |
| | | CTC | L | CCC | P | CAC | H | CGC | R | C | |
| | | CTA | L | CCA | P | CAA | Q | CGA | R | A | |
| | | CTG | L | CCG | P | CAG | Q | CGG | R | G | |
| | A | ATT | I | ACT | T | AAT | N | AGT | S | T | |
| | | ATC | I | ACC | T | AAC | N | AGC | S | C | |
| | | ATA | I | ACA | T | AAA | K | AGA | R | A | |
| | | ATG | M | ACG | T | AAG | K | AGG | R | G | |
| | G | GTT | V | GCT | A | GAT | D | GGT | G | T | |
| | | GTC | V | GCC | A | GAC | D | GGC | G | C | |
| | | GTA | V | GCA | A | GAA | E | GGA | G | A | |
| | | GTC | V | GCG | A | GAG | E | GGG | G | G | |

## Key:

G – Glycine (Gly)

P – Proline (Pro)

A – Alanine (Ala)

V – Valine (Val)

L – Leucine (Leu)

I – Isoleucine (Ile)

M – Methionine (Met)

C – Cysteine (Cys)

F – Phenylalanine (Phe)

Y – Tyrosine (Tyr)

W – Tryptophan (Trp)

H – Histidine (His)

K – Lysine (Lys)

R – Arginine (Arg)

Q – Glutamine (Gln)

N – Asparagine (Asn)

E – Glutamic Acid (Glu)

D – Aspartic Acid (Asp)

S – Serine (Ser)

T – Threonine (Thr)

# Contents

# List of Figures

## Chapter Four - A variant in the *EIF3H* promoter influences colorectal cancer risk, patient survival and response to treatment

## Chapter Five - Multiple colorectal cancer susceptibility alleles affect patient survival and rs9929218 in E-cadherin predicts response to cetuximab

## Chapter Six - Identifying biallelic high penetrance CRC susceptibility alleles in the BER, MMR and OxDR pathways

## Chapter Seven - General Discussion

## List of Tables

# Description of contributions to this thesis

**Chapter 03** – FFPE and blood samples of COIN patients had previously been collected prior to the commencement of the PhD. Processing of the vast majority (60-80%) of tumour samples was carried out by Christopher Smith (CS), with further support from Michelle James, Rebecca Harris (RH) and Shelley Idziaszczyk (SI). This included microtome sectioning, H&E staining, tumour extraction (LCM and macrodissection) and DNA extraction.

All PCRs for pyrosequencing were carried out by CS. PCRs used in Sequenom were carried out by the Lambrecht's group (with the support of CS). All Pyrosequencing was carried out by CS. All Sequenom was carried out by the Lambrechts group with the support of CS. Results for Pyrosequencing were independently viewed by CS and RH, and compared. Results for Sequenom were viewed by CS and compared against calls made by the Sequenom analysis software. Pyrosequencing and Sequenom genotypes were compared by CS. Sanger sequencing of select COIN samples was carried out by CS and RH. All MSI analyses (from PCR to BAT25/26 analysis) were carried out by CS.

Clinical data relating to COIN patient drug response was collected and analysed by the medical research council clinical trials unit (MRC CTU). Statistical analyses were carried out through a pre-defined statistical plan defined by the CTU. Overlap between mutant subgroups was studied by CS. Interpretation of clinical data with relation to mutation status was independently carried out by the CTU and CS/Jeremy Cheadle (JPC)

**Chapter 04** – DNA extraction from COIN patient blood was primarily carried out by SI with the support of CS. Preparation of COIN and UKBS DNA samples was carried out by SI/CS. Genotyping, through the Illumina GoldenGate platform, was arranged through JPC/CS and James Colley (JC). Statistical interpretation of genotyping data was carried out by Valentina Moskvina (VM). Association with patient survival were carried out by David Fisher (DF) of the MRC CTU. All downstream functional and statistical analyses described in chapter 04 were carried out by CS.

**Chapter 05** – Genotyping, through Geneservice and Kbiosciences, was arranged by CS. Trial design was carried out by CS and JPC. Associations with patient survival, response and toxicity were carried out by DF. Interpretation of statistical data was carried out by CS and JPC.

**Chapter 06** – All experimental and statistical work described in this chapter was carried out by CS with the support of the Cheadle laboratory.

# Chapter One - General Introduction

## 1.1 Colorectal Cancer

Colorectal cancer (CRC) is one of the most common cancers worldwide, with over one million new cases diagnosed each year (with incidence looking set to rise with lifestyles in Asian and African populations becoming progressively 'westernised'). It is the third most common cancer in the UK with around 106 new cases diagnosed each day and with ~38,500 registered cases in 2007 (Cancer Research UK, UK incidence statistics, 2010). Mortality rates have declined in recent years (Edwards *et al.* 2010) thanks to increased screening, reduction of risk factors and improved treatment. Despite these improvements, the disease-specific mortality rate is still approximately 33% in the developed world (Wolpin *et al.* 2008)

Environmental and genetic factors contribute to the development of CRC and either can be the predominant cause of the disease. They can also interact, as demonstrated by a genetic predisposition that results in an increased susceptibility to particular carcinogens in the diet (Menko, 1993). Environmental factors include age - 90% of CRC cases are over the age of 50. Also, life-style choices such as exercise, smoking habit, alcohol intake, and poor diet can contribute to risk. Kolonel and colleagues (2004) highlighted the importance of these environmental factors when studying Japanese immigrants to Hawaii. Here the incidence of CRC in even the first generation of migrants was found to reach increased rates and was likely a result of changes in diet. Diets low in fiber and high in red meat, fat and carbohydrates (Chan and Giovannucci, 2010) have been associated with increased risk (World Cancer Research Fund, 2007).

## 1.2 Heritable CRC

As presented in Figure 1.1, the majority of CRCs arise sporadically. However, there is a considerable heritable component including a number of CRC syndromes. It is estimated that the familial category has a two- to three-fold greater risk than the general population (Cunningham *et al.* 2010). Hereditary CRC syndromes have important public health implications. Correct diagnosis of an inherited syndrome, on the basis of a family pedigree or through molecular characterisation, is essential in

disease management and clinical decision making for further testing of family members (Lynch *et al.* 2007).

## 1.2.1 High penetrance mechanisms of inheritance

Inherited factors are thought to play a significant role in up to one third of CRCs (Lichenstein *et al.* 2000), but only a minority of these can be accounted for by established CRC predisposition genes (Kinzler and Vogelstein 1996). <6% of cases carry high-penetrance germline mutations (Aaltonen *et al.* 2007) such as those found in the adenomatous polyposis coli (*APC*) gene, which cause familial adenomatous polyposis (FAP) (Fearnhead et al. 2001). FAP, and other inherited conditions arising through high penetrance mutations, are described below;

## 1.2.1.1 FAP

FAP is an autosomal dominant inherited disorder that affects about 1 in 7000-8000 individuals (Cunningham *et al.* 2010; Half *et al.* 2009), manifesting equally in both sexes. Individuals with FAP are predisposed to hundreds to thousands of adenomatous polyps within the colon and rectum (Wallis and MacDonald, 1996). Whilst these polyps that develop in the second to third decade of life (Kinzler and Vogelstein, 1996) are not cancerous, their presence greatly increases the likelihood of the individual developing CRC – left untreated, one or more will progress to carcinoma often by the fourth decade of life (Hodgson and Maher, 1999; Kinzler and Vogelstein, 1996) and almost all patients will develop CRC if not recognised and treated at an early stage. A phenotypically 'less-severe' form of FAP also exists, termed attenuated-FAP (AFAP). This is characterised by a reduced adenoma burden and a later age of adenoma (and thus cancer) development.
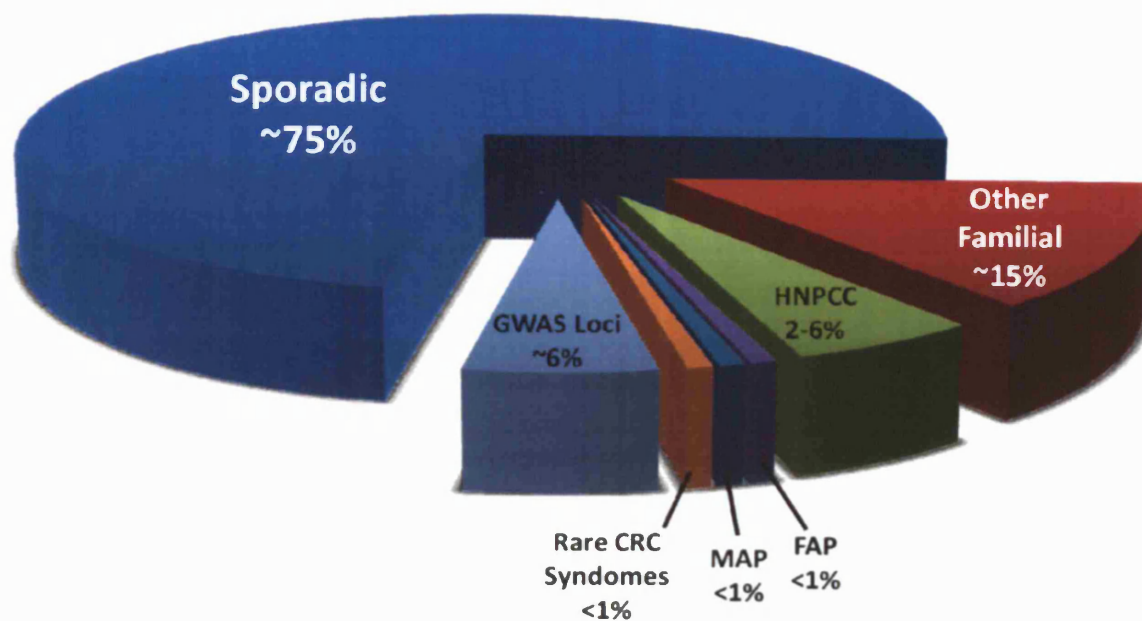
**Figure 1.1** – Pie chart depicting the frequencies of the different forms of CRC. Seventy-five percent of CRCs are sporadic in nature, leaving ~25% with some form of inherited component. Of the known disorders, HNPCC accounts for the most cases (2-6%, of CRCs) with other conditions such as MAP and FAP accounting just under 1% each. The GWAS loci individually confer very modest contributions to risk but collectively account for ~6% of CRC risk. About 15% of cases are due to an unknown familial component but this may be accounted for by epistatic interactions between the other known factors.

### 1.2.1.1.1 The *APC* gene

FAP arises through mutations of the *APC* tumour suppressor gene. The gene consists of at least 21 exons (Santoro and Groden, 1997) spanning a region of >100 kilobase pairs (kbp). The last exon comprises the majority of the coding sequence (~77%) and is the location of most germline and somatic mutations.

More than 300 different mutations of APC can contribute to the FAP phenotype and these come in the form of insertions, deletions and nonsense mutations that result in truncation of the APC protein. The majority of germline mutations result in the synthesis of a C-terminally truncated protein (Béroud and Soussi, 1996; Laken *et al.* 1999; Powell *et al.* 1992) or show selection against β-catenin regulating domains (Polakis, 2000). Large deletions that cover whole exons, or the whole gene itself, have been observed and contribute to ~5% of FAP cases (Sieber *et al.* 2002; Michils *et al.* 2005). Somatic mutations provide the 'second hit' that is required for tumour formation and are found to cluster within a central region of the gene. This central region (known as the mutation cluster region or MCR) lies between codons 1281 and 1580 and greater than 60% of somatic *APC* mutations occur here (Cheadle *et al.* 2002; De Rosa *et al.* 2003).

### 1.2.1.1.2 Wnt Signalling and cancer

The Wnt pathway is critical in controlling important processes such as cell proliferation and differentiation. APC is a key part of this pathway, functioning as part of a destruction complex that targets β-catenin for degradation. Wnt signals influence the stability of this destruction complex that also includes casein kinase 1α (CK1), glycogen synthase kinase 3 (GSK3) and the scaffolding protein Axin. The overall process is depicted in Figure 1.2. The regulation of Wnt signalling is a critical process and its loss is associated with cancer initiation - often through the loss of activity of genes encoding members of the destruction complex or activating mutations of β-catenin (Sparks *et al.* 1998) itself. Abnormal levels of β-catenin in-turn lead to increased expression of Wnt target genes which contribute to the initiation of tumourigenesis.
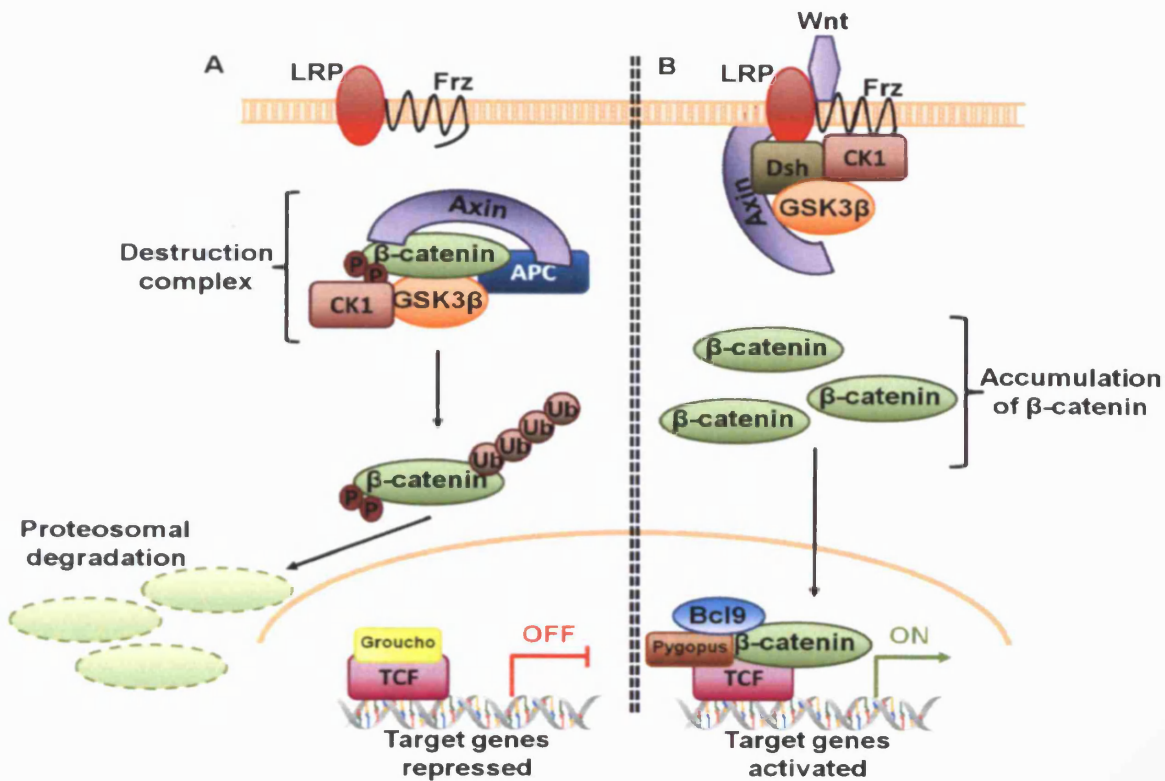
**Figure 1.2** - Wnt signalling in colonic epithelial cells. (A) In the absence of Wnt, β-catenin is recruited to the destruction complex (that includes APC, CK1, GSK3β and Axin). APC and Axin (which provides the scaffold for this complex) physically interact with β-catenin, while CK1 and GSK3β phosphorylate the N-terminus of the protein (Sakanaka *et al.* 1999). It is subsequently ubiquitinated by an E3 ubiquitin ligase called β-TrCP and is proteasomally degraded. This restriction on the levels of cytoplasmic β-catenin enables the interaction of the transcriptional repressor Groucho and the TCF/LEF machinery and ensures that Wnt target gene transcription is repressed (Daniels and Weis, 2005). (B) When Wnt binds to the transmembrane receptor Frz and LRP5/6 co-receptors, Axin translocates to the cell membrane and interacts with LRP5/6 and Disheveled is phosphorylated by CK1, preventing GSK3β from phosphorylating β-catenin (Clevers, 2006). Intracellular β-catenin is able to accumulate and translocate to the nucleus where it binds with TCF/LEF transcription factors and other components (Pygopus and Bcl9), thus activating the transcription of Wnt target genes. These include the polyamine ornithine decarboxylase (ODC) proto-oncogene (Bello-Fernandez *et al.* 1993), *CCND1* (Tetsu and McCormick, 1999) and *EphB* (Batlle *et al.* 2002). Abnormal expression of these genes can contribute to the neoplastic transformation of the cell.

## 1.2.1.2 HNPCC

Accounting for 2-6% of all CRC cases, HNPCC is the most common hereditary form of CRC (Lynch *et al.* 2006; Hampel *et al.* 2008). It has an autosomal dominant inheritance pattern (Potter, 1999) that manifests at an average age of 45 years (compared to 69 years for sporadic CRC in the general population) (Lynch *et al.* 2009). Tumourigenesis shows a predilection towards proximal colon growth, with accelerated development from adenoma to carcinoma (within 2-3 years in HNPCC patients as compared to 8-10 years in the general population). Despite this, adenomas are rarely present in large numbers (exceptionally more than 10) and florid polyposis is not observed (Maher *et al.* 1991). Amongst HNPCC sufferers, there is a high risk of additional CRCs with ~30% of those that undergo surgery developing a second primary tumour within 10 years of resection (Lynch *et al.* 2009). Sufferers show longer survival periods than other forms of CRC. HNPCC is caused by germline mutations in genes involved in the mismatch repair (MMR) pathway – mutations in the *MSH2* (Fishel *et al.* 1993), *MLH1* (Lindblom *et al.* 1993), *MSH6* (Miyaki *et al.* 1997), *MLH3* (Wu *et al.* 2001) or *PMS2* (Nicolaides *et al.* 1994) have been linked to this condition.

## 1.2.1.2.1 DNA mismatch repair

DNA MMR is a cellular repair system that repairs mismatches and insertion-deletion loops. During replication, slippage can lead to misalignment of the template and synthesis strands, in-turn giving rise to heteroduplex DNA molecules. Microsatellites are short repetitive DNA motifs of 1-6 nucleotides in length that are repeated 10-60 times (Beckmann and Weber, 1992). They are polymorphic in length among individuals but are unique and uniform in length in every tissue within an individual (Boland and Goel, 2010) and are abundant throughout the genome though are more common in non-coding regions. Mismatches can contribute to the microsatellite instability (MSI) phenotype that is observed in 15% of CRCs. The MMR system repairs the section of the synthesis strand containing the error in a process summarised in Figure 1.3.
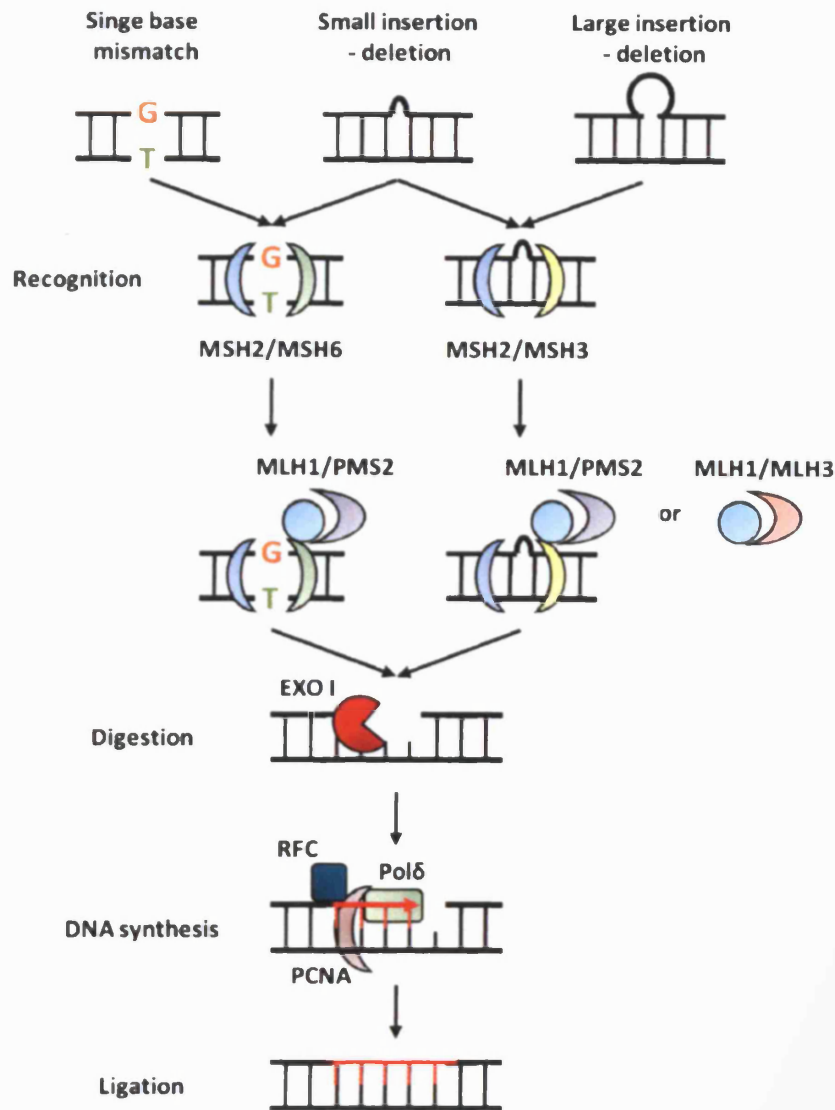
**Figure 1.3** – Diagram depicting the human MMR system. Initially the repair machinery must recognise base-base mismatches or insertion-deletion loops. In humans this is carried out by two separate complexes; MSH2/MSH6 which recognise mismatches and small insertion-deletion loops, and MSH2/MSH3 which recognise all sizes of insertion-deletion loops (Culligan *et al.* 2000). After recognition, a MLH1/PMS2 heterodimer is recruited to the damage site and docks to a MSH2 heterodimer (Jascur and Boland, 2006; Plotz *et al.* 2006). The complex of MLH1/MLH3 has a largely unclear function but may participate in the repair of a subset of mismatches recognised by the MSH2/MSH3 complex. Repair continues with the exonuclease 1 (EXO1) mediated degradation of DNA until the lesion is removed. Finally, the resulting excision tract is filled in by polymerase δ (Polδ) and the correct nucleotide will be inserted in the original mismatch position, and the remaining nick is sealed by ligation.

## 1.2.1.2.2 Germline mutations in MMR genes

Mutations within the *MLH1* and *MSH2* genes account for >95% of HNPCC cases (Peltomäki, 2005). The remaining mutations are distributed amongst the other MMR genes (Papp *et al.* 2007). The majority of germline mutations are nonsense or frameshift mutations, though missense, splice-site variants and small- and large-deletions and insertions have been observed (Kurzawski *et al.* 2006). Germline mutations do not cluster at 'hot-spots' within the MMR genes, and the majority of the mutations have only been reported once.

## 1.2.1.2.3 Somatic mutations in HNPCC

The MMR genes act as tumour suppressor genes – both alleles are required to be inactivated to contribute to tumourigenesis since cells heterozygous for a mutation function normally (Parsons *et al.* 1993). Mutational inactivation of the MMR machinery leads to deficiencies in the MMR system that causes the MSI phenotype.

## 1.2.1.2.4 MSI

MSI is observed in the majority of tumours in HNPCC cases, but is only detected in ~10-15% of sporadic CRCs (Thibodeau *et al.* 1993; Ionov *et al.* 1993). Sporadic CRCs displaying MSI arise through a process that involves the CpG island methylator phenotype (CIMP) (Toyota *et al.* 1999). CpG islands are clusters of cytosine-guanine nucleotides that are often found within the promoter regions of coding elements (Issa, 2004). These genes can be silenced through a mechanism in which the cytosine residue of this di-nucleotide becomes methylated through the action of methyltransferases. One such target of this 'hypermethylation' is the *MLH1* gene (Kuismanen *et al.* 2000). CRCs arising through the CIMP mechanism are distinct from HNPCC cases (Nagasaka *et al.* 2008; Boland *et al.* 2009).

MMR deficiency leads to an increase in the rate of point mutation accumulation across the genome of a cell, with a bias towards mutations in repetitive sequences. When microsatellites lying within coding regions of regulatory genes become disrupted, MSI can contribute towards the development of cancer.

## 1.2.1.3 MAP

The discovery of this condition came about through the study of a British family in which three siblings were affected by multiple colorectal adenomas and carcinomas but none had an inherited mutation of APC (Al-Tassan *et al.* 2002). Equal expression of both APC alleles was observed and there was no evidence for defective MMR. Upon analysis of the somatic status of APC in the patients tumours, an over-representation of G:C→T:A transversions was observed. The number of transversions was compared against those in 811 colorectal carcinomas, and it was found that the high-occurrence within the three siblings was statistically significant (p=$10^{-12}$). Based on the knowledge that the *E.coli* MutY, MutM and MutT glycosylases counteract and repair damage resulting from the mispairing of adenine with 8oxoG, that result in such G:C→T:A transversions (Steenken *et al.* 1997), Al-Tassan and colleagues screened the human orthologues of these genes; *MUTYH, OGG1,* and *MTH1* respectively, for pathogenic mutations. Germline bi-allelic mutations of *MUTYH* were observed in all three affected siblings. These bi-allelic mutations were found at Y165C and G382D (though the current nomenclature refers to these as Y179C and G396D respectively). The occurrence of these variants in a heterozygous, non-biallelic state within phenotypically normal controls suggested that their mutagenic effect was recessive. Further carriers of bi-allelic MUTYH mutations were found within the adenoma and polyposis setting by Jones (2002) and Sieber (2003). Functional analysis of the variants within the *E.coli* orthologues revealed that their presence functionally compromises the protein, with the ability to remove adenine being severely depleted in Y165C mutant- and 85% decreased within G382D mutant-carriers.

MAP is an autosomal recessive inherited condition that accounts for ~1% of all CRCs (Fleishmann *et al.* 2004). The colorectal phenotype of MAP resembles AFAP (<100 polyps) and 'mild' FAP (hundreds to thousands of polyps), but not severe FAP.

## 1.2.1.3.1 MUTYH function

MUTYH is a DNA glycosylase that contributes to the maintenance of genomic stability through its role in the base excision repair (BER) of oxidative DNA damage. In humans, there are approximately $10^4$ lesions generated per cell, per day as a result of oxidative damage to DNA (Ames and Gold, 1991). The main causes of this

9

oxidative damage are reactive oxygen species (ROS) which include hydroxyl radicals, superoxide and hydrogen peroxide. ROS are by-products of cellular metabolism as well as environmental exposures and have been linked to the initiation and progression of cancer, as well as degenerative diseases and ageing (Klaunig and Kamendulis, 2004). Types of damage resulting from ROS include strand-breaks or lesions of individual bases. Nucleotides are susceptible to oxidative damage, especially guanine due to its low oxidative potential (Steenken et al. 1997). A number of intracellular mechanisms exist to monitor and respond to this damage - extensive damage may result in cell death while limited damage can be repaired with little consequence to the cell. However, mutations can result from replication if the modified bases are not repaired (Lu et al. 2001).

## 1.2.1.3.2 BER

BER is the main mechanism that protects the cell against oxidative DNA damage (Hazra et al. 2007) though it can also repair lesions that result from methylation, deamination and hydroxylation (Hoeijmakers 2001). The overall BER process is presented in Figure 1.4.

7, 8-dihydro-8-oxoguanine (8oxoG) is the most stable product of oxidative DNA damage (Ames and Gold, 1991). 8oxoG can functionally mimic the thymine base, and thus readily mispairs with adenine leading to G:C→T:A transversions. It is estimated that there are 0.3-4.2 8oxoG bases per $10^6$ guanine bases (Collins et al. 2004) and high concentrations have been detected in breast, lung, and kidney cancers (Malins and Haimanot, 1994; Olinski et al. 1992; Jaruga et al. 1994; Okamoto et al. 1994). Thus it is essential that the cell can recognise and remove the lesion. Cytosine can also be 'misincorporated' opposite 8oxoG but this is observed 5-200 times less frequently than the mis-incorporation of adenine (Shibutani et al. 1991). E.coli was first used to study the defence against damage resulting from spontaneous 8oxoG. Three key enzymes were identified; MutT, MutM and MutY which function as a group of BER glycosylases (Michaels and Miller, 1992). Since then, human homologues of these key glycosylases have been identified as MTH1/NUDT1 (MutT; Sakumi et al. 1993), OGG1 (MutM; Roldàn-Arjona et al. 1997; Arai et al. 1997) and MUTYH (MutY; Slupska et al. 1996) which are also critical in 8oxoG repair, as described in Figure 1.5)
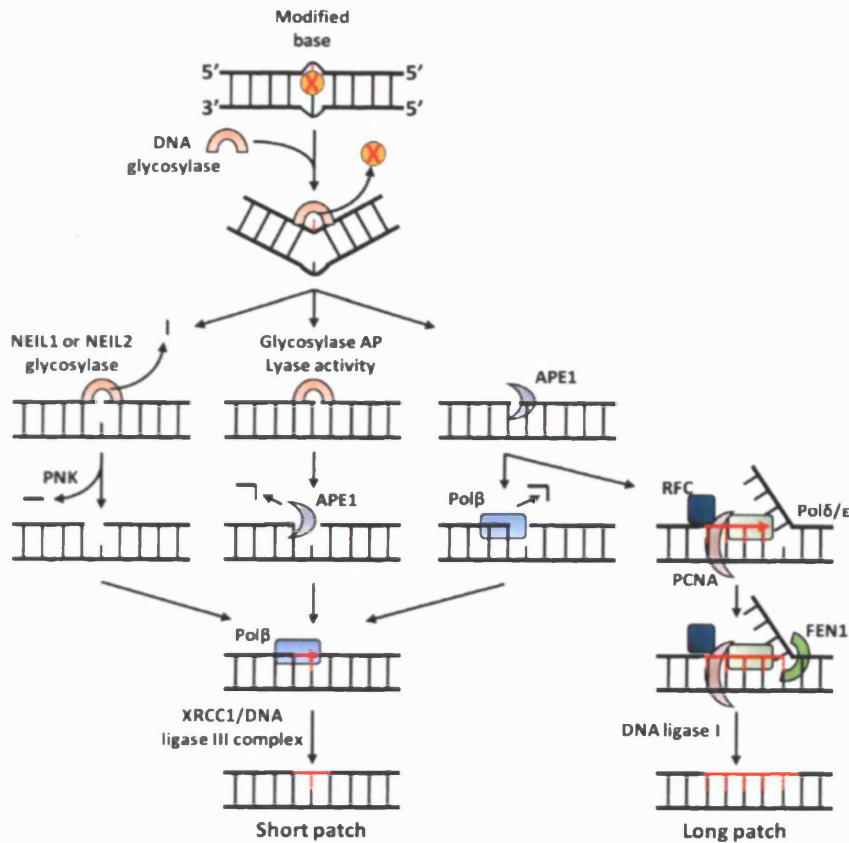
**Figure 1.4** – Diagrammatic representation of the BER process. Here a DNA glycosylase will recognise a damaged base and 'flip' it out of the helix by DNA backbone compression. This, in-turn, allows the DNA glycosylase to accommodate the base into its active site and cleave it from the sugar-phosphate backbone. The resulting apurinic/apyrimidinic (AP) site (which can also occur spontaneously by hydrolysis) can also be cleaved 3' to this site by glycosylases that specifically target AP sites. The APE1 endonuclease then processes the DNA in preparation for DNA synthesis. At this stage, the repair pathway diverges into short-patch or long-patch BER (Wilson and Bohr, 2007), with the former being the more dominant mode in mammals. In short-patch BER, DNA polymerase β performs a one-nucleotide gap-filling reaction and removes the 5' terminal baseless sugar residue via its lyase activity. The remaining nick is then sealed by an XRCC1 (X-ray repair cross complementing 1)/DNA ligase complex. In the long-patch BER pathway, DNA polymerase β incorporates the first nucleotide and polymerase δ/ε carry out elongation. Following this, a repair patch of two or more nucleotides is synthesised from the damaged site, facilitated by proliferating cell nuclear antigen (PCNA) and replication factor C (RFC). The displaced DNA flap is then excised by flap endonuclease 1 (FEN1) before a final ligation reaction is carried out by DNA ligase 1. The above mechanisms occur across the genome, with the divergence into the dominant short-patch or minor long-patch repair pathways dependent upon which DNA glycosylase is recruited to the damage site, as well as the type of damage present and the current cell cycle phase.
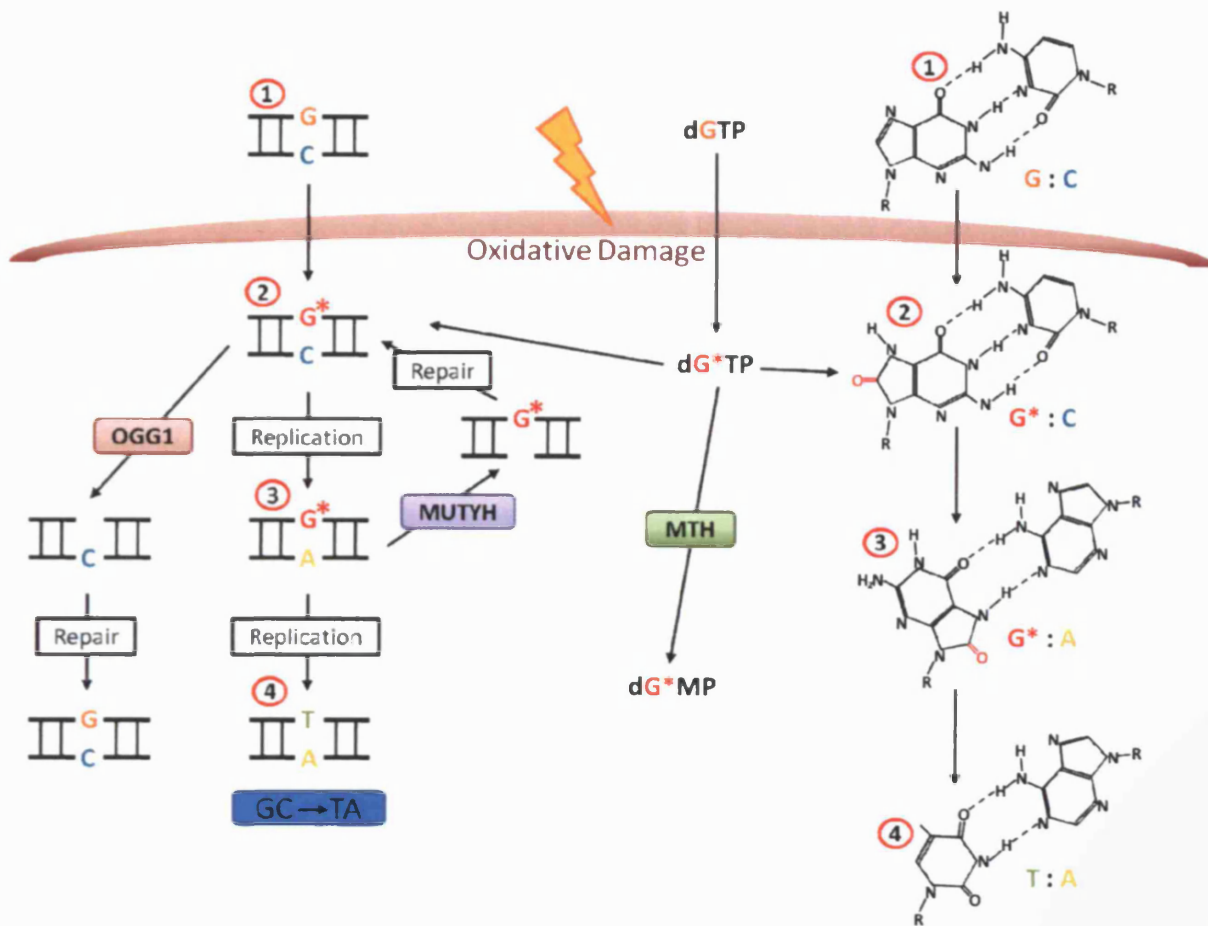
**Figure 1.5** – The roles of human MUTYH, OGG1 and MTH1 in the repair of the 8oxoG (shown as a G*) lesion. Presented are the structures and binding patterns of the normal and damaged bases during different stages of the repair process. The major enzyme for the repair of 8oxoG is considered to be OGG1 (Hirano, 2008). OGG1 (scarlet box) recognises 8oxoG when it is paired with cytosine (as shown by point 2) and excises it. The repair site can subsequently be replaced with the correct base, G. There are reports that OGG1 may recognise and act upon other lesions, albeit less effectively (Bjørås *et al.* 1997). MUTYH (purple box) has glycosylase activity against a number of DNA substrates including the mispairs; A:G, 2-hydroxyadenin (2-OH-A; aka isoguanine):8oxoG, A:8oxoG, 2-OH-A:G, and 2-OH-A:A (Ohtsubo *et al.* 2000). Here, it excises the mispaired adenine (as shown by point 3), allowing OGG1 a second opportunity to excise 8oxoG from the regenerated 8oxoG:C mispair. MTH1 (green box) is an oxidised purine nucleoside triphosphatase that efficiently hydrolyzes oxidized dGTP, GTP, dATP and ATP in the nucleotide pool, preventing their incorporation into the nascent strand during DNA replication (Nakabeppu *et al.* 2001). In the case of 8-oxodGTP (dG*TP above), MTH1 degrades it to 8-oxodGMP (dG*MP) which, upon further dephosphorylation, is either excreted in the urine or recycled (Cooke and Evans, 2007; Hah *et al.* 2007).

12

When mutated, MUTYH has a significantly lower capacity for binding and repair of A:8oxoG and 8-oxoadenine:G mispairs (Parker *et al.* 2005). This defective repair of lesions defines the MAP condition. Little is known about the contribution of OGG1 towards cancers - given the importance MUTYH in predisposition to CRC, it was thought that OGG1 may play a similar role based its common role in protection against the 8oxoG lesion but little evidence for this has been found. Possibly the strongest candidate for an inactivating or disrupting mutation that contributes to cancer is the common Ser326Cys variant (Kohno *et al.* 1998; Dherin *et al.* 1999), though other polymorphisms, with as yet unknown consequences, could contribute (Sidorenko *et al.* 2009). A screen of 167 unrelated patients with multiple colorectal adenomas for inherited variants failed to identify any biallelic combinations of rare alleles in *MTH1* and *OGG1* (as well as other BER enzymes *NEIL1*, *NEIL2*, *NEIL3* and *NTH1*) (Dallosso *et al.* 2007).

## 1.2.2 Low penetrance mechanisms of inheritance

Beyond the high penetrance mechanisms of inheritance, it has been proposed that a substantial proportion of the remaining inherited predisposition to CRC is caused by either rare (Fearnhead *et al.* 2005) or common low penetrance variants.

## 1.2.2.1 'Common-disease rare variant' hypothesis

In support of the 'rare variant hypothesis', Azzopardi and colleagues (2008) have shown that individually rare, but collectively common, inherited non-synonymous variants in *APC* play a significant role in multi-factorial inherited predisposition to colorectal adenomas (CRAs). The non-synonymous variants I1307K and E1317Q had previously been suggested by some studies to predispose to CRAs and CRCs (Frayling *et al.* 1998; Lamlum *et al.* 2000; Gryfe *et al.* 1999; Hahnloser *et al.* 2003), while others had not supported these observations (Gismondi *et al.* 2002; Popat *et al.* 2000). Azzopardi and colleagues confirmed a role for E1317Q as a low penetrance variant for predisposition to CRAs by virtue of its significant over-representation in 'non-FAP, non-MAP' patients as compared to controls. Other investigators have also reported that rare variants in the Wnt signalling genes *AXIN1* and *CTNNB1* and the MMR genes *MLH1* and *MSH2* contribute to colorectal tumourigenesis (Fearnhead *et al.* 2004)

## 1.2.2.2 'Common-disease common variant' hypothesis

In the 'common-disease common-variant' model for CRC, the risk associated with individual variants is small; however, they make a significant contribution to the overall disease burden by virtue of their high frequencies in the population. Moreover, by acting in concert with each other, they have the potential to significantly affect an individual's risk of developing CRC. Recent years have seen the identification of such variants through the use of genome wide association studies (GWAS)

## 1.2.2.3 GWAS of CRC

GWAS have become possible thanks to a number of key factors. Firstly, the collection of large and well characterised case and control cohorts have provided researchers with study groups large enough to overcome the power issues faced in large scale statistical analyses. Secondly the completion of the HapMap project (the international HapMap consortium, 2003) has allowed the selection of tag single nucleotide polymorphisms (SNPs) that capture much of the variation of the genome. Finally, technological advances in platforms used for large-scale genotyping have allowed realistic costs and time-scales for researchers when screening large numbers of variants across many thousands of samples.

Several studies have conducted multistage GWA and, to-date identified 14 low penetrance susceptibility loci for CRC risk (Table 1.1). The emergence of these variants has yielded advances in the understanding of the role played by common variation in CRC. It is estimated that each locus contributes less than 1% to familial risk of CRC (Houlston *et al.* 2008), in an independent fashion. Using an additive model, and considering that an individual will likely carry more than one risk variant, the 14 loci collectively account for ~6% of the excess familial risk (Figure 1.1). Furthermore, they allow the possibility of developing multi-locus prediction models of genetic risk that could be combined with other known risk factors (Tenesa and Dunlop, 2009).

| Locus | Association SNP/s | Gene | Reference |
|---|---|---|---|
| 1q41* | rs6691170, rs6687758 | DUSP10 | Houlston et al. 2010 |
| 3q26 | rs10936599 | MYNN | Houlston et al. 2010 |
| 8q23 | rs16892766 | EIF3H | Tomlinson et al. 2008 |
| 8q24 | rs6983267 | - | Tomlinson et al. 2007 |
| 10p14 | rs10795668 | - | Tomlinson et al. 2008 |
| 11q23 | rs3802842 | - | Tenesa et al. 2008 |
| 12q13* | rs11169552, rs7136702 | LARP4 | Houlston et al. 2010 |
| 14q22 | rs4444235 | BMP4 | Houlston et al. 2008 |
| 15q13 | rs4779584 | - | Jaeger et al. 2008 |
| 16q22 | rs9929218 | CDH1 | Houlston et al. 2008 |
| 18q21 | rs4939827 | SMAD7 | Broderick et al. 2007 |
| 19q13 | rs10411210 | RHPN2 | Houlston et al. 2008 |
| 20p12 | rs961253 | - | Houlston et al. 2008 |
| 20q13 | rs4925386 | LAMA5 | Houlston et al. 2010 |

**Table 1.1** – Table highlighting the 14 distinct loci associated with CRC identified through GWAS. Some SNPs map within or near to genes though in most cases, it is unclear as to whether the SNP actually interacts with the gene. *Two SNPs associate with risk at these loci.

## 1.2.2.3.1 8q24 and the risk mechanism at this locus

The first CRC risk locus to be identified through GWAS was 8q24 (Zanke *et al.* 2007). A multistage association of 7,480 CRC cases and 7,779 controls highlighted the 8q24 locus (especially in a region of linkage disequilibrium (LD) that houses SNPs rs6983267 and rs10505477) as being linked to CRC risk. A separate analysis (Tomlinson *et al.* 2007), involving a series of case/control analyses identified rs6983267 as the most strongly associated SNP at this locus – this association also included CRC adenoma cases, suggesting that the risk at this locus contributes towards tumour initiation. Attention had previously been drawn to the 8q24 region in previous studies of CRC - the *MYC* oncogene, which resides at this locus, was amplified in 32% of CRC patients in a study by Augenlicht and colleagues (1997). In addition, the tyrosine phosphatise PRL-3 at 8q24 is amplified in CRC metastasis (Mollevi *et al.* 2008). Many other studies of various populations have validated the association at this locus (Gruber *et al.* 2007; Poynter *et al.* 2007; Li *et al.* 2008; Tuupanen *et al.* 2008; Berndt *et al.* 2008; Curtin *et al.* 2009; Hutter *et al.* 2010). As well as CRC, the 8q24 locus has been linked with other cancers including prostate, bladder and breast (Haiman *et al.* 2007; Kiemeney *et al.* 2008; Easton *et al.* 2007). Accounting for the various associated cancers, there appear to be at least 5 different susceptibility loci at 8q24.

Recent findings have started to shed light on the functional effect rs6983267 might be having (Figure 1.6). The risk allele (G) of rs6983267 enhances binding of TCF4 which in-turn regulates *MYC,* a target of the Wnt pathway (Tuupanen *et al.* 2009). Further evidence for this interaction was found by Pomerantz and colleagues (2009). Using CRC cell lines, the region was confirmed as being a binding site for *TCF4,* with the G allele showing greater binding potential than the T allele. The chromosome conformation capture (3C) technique revealed that, despite the large genomic distance between the risk region and *MYC* gene, a physical interaction between the two exists – specifically including the first half of *MYC* and its promoter. Wright and colleagues (2010) confirmed this by revealing that TCF4 physically interacts with *MYC* through the formation of chromosome loops regardless of which allele is present at rs6983267. The presence of the G allele, however, correlates with increased MYC expression (~2-fold higher than the T allele). Furthermore Sotelo and colleagues (2010) demonstrated significantly varied enhancer activity depending

upon which allele of rs6983267 was present – contradictory to the above findings, the 'non-risk' T allele correlated with greater *MYC* activity than did the risk G allele. They suggest that rs6983267 acts to regulate *MYC* and its oncogenic effects through union of intrinsic thresholds that regulate cellular proliferation, and tumour suppression. MYC expression, when raised above a certain level, will trigger cellular apoptosis (Jiang *et al.* 2003) but low levels of MYC will drive cellular proliferation unchecked.

## 1.2.2.3.2 18q21 and the risk mechanism at this locus

Shortly after the identification of 8q24 as a CRC risk locus, 18q21 was highlighted as housing a variant, rs4939827, that also associates with risk (P=1.0x10$^{-12}$) (Broderick *et al.* 2007). This variant maps to the third intron of the *SMAD7* gene (mothers against decapentaplegic homolog 7) which is an intracellular antagonist of TGFß signalling that acts by binding the active receptor complex and blocking downstream signalling. Since the TGFß signalling pathway is involved in many critical cellular processes including cell growth, apoptosis, cellular differentiation, adhesion and migration, and immune response, *SMAD7* seemed the most obvious functional target for the risk associated allele of rs4939827. The 18q21 locus accounts for ~0.8% of excess familial risk of CRC. The modest risk at this locus was further replicated in the Tenesa's 2008 study (P=7.8x10$^{-28}$, OR=1.2) as well as through the use of the case/control cohorts of Curtin and colleagues (2009) (P=0.007, OR=0.77) and Slattery and colleagues (P<0.01, OR=0.77) (2010).
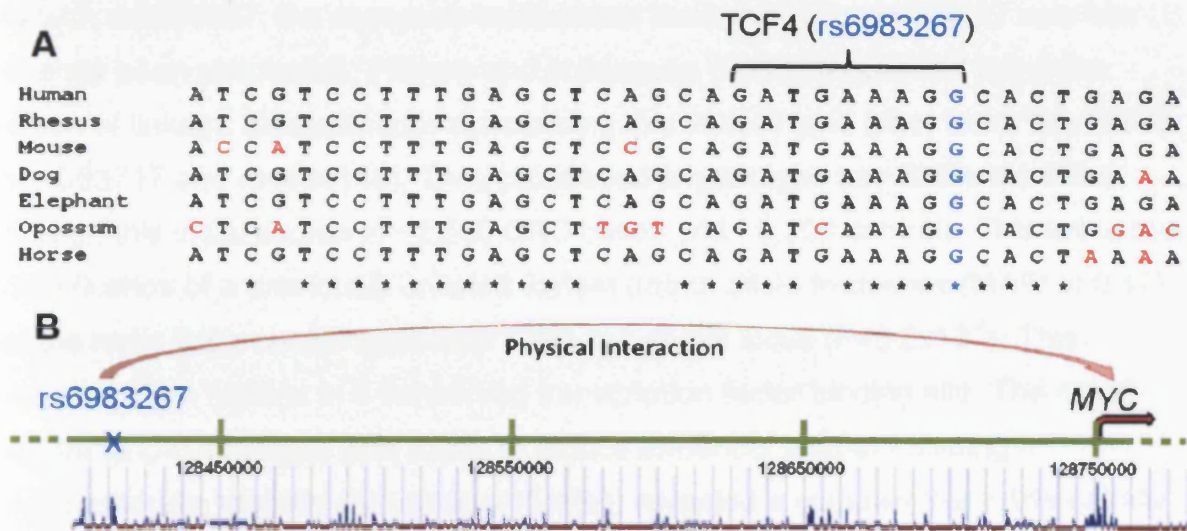
**Figure 1.6** – Suggested mechanism through which rs6983267 (in blue) contributes to CRC risk despite being >300kb telomeric to the *MYC* gene. Tuupanen and colleagues (2009) initially identified a potential role for the SNP as part of an enhancer element using the 'enhancer element locator' which identifies potential distal enhancer elements through the alignment of conserved mammalian sequences (A, Conserved bases are black and represented with an *). This highlighted the immediate region surrounding the SNP as a binding site for the Wnt regulated transcription factor *TCF4*. Further studies have replicated this, and shown that this region physically interacts with the *MYC* gene (B).

As with rs6983267, the causative mechanism through which rs4939827 acts has recently been uncovered. Pittman and colleagues (2009) sequenced the entire region of linkage disequilibrium surrounding this variant (and other lower risk SNPs rs12953717 and rs4464148). They proceeded to genotype any SNPs identified through this initial screen in ~2,500 CRC cases and ~2,500 controls. This led to the identification of a previously unlisted variant (minor allele frequency (MAF) of 0.47) as the most likely causal variant for CRC risk at this locus (P=5.2x10$^{-6}$). This polymorphism resides in a conserved transcription factor binding site. The novel variant (a C>G change) was found to reduce enhancer element binding – electrophoretic mobility shift assays (EMSA) revealed a complex that preferentially forms when the C allele is present. In addition, the G allele correlated with significantly reduced expression of green fluorescent protein (GFP) in construct experiments. Finally, the authors highlight the contributory effect towards CRC through SMAD7 expression analysis of cancer and adenoma cases, with significantly lower expression observed in cancers.

### 1.2.2.3.3 The remaining GWAS loci

Like those at 8q24 and 18q21, most of the remaining risk variants map to regions of seemingly little genomic consequence (i.e. intronic or variants located in 'gene-deserts') and as yet the functional consequences behind them have not been identified. The occurrence of GWAS SNPs (for other conditions as well as CRC) in 'gene-desert' genomic locations poses a large problem for future research. A recent meta-analysis revealed that roughly 40% of ~1200 disease associated SNPs lie in LD blocks tagging no coding elements (Visel *et al.* 2009).

Despite the productive output of GWAS in recent years, it remains likely that, outside of the top 20-50 most significant SNPs for a particular condition, other variants will not be detected without exponential increases in sample sizes analysed, due to limitations in power.

### 1.2.2.4 Pathways of CRC susceptibility

GWAS have identified several components of the TGFβ pathway as carriers of risk alleles for CRC. These include SMAD7 (18q21), GREM1 (15q13), the bone morphogenetic protein genes BMP2 and BMP4 (20p12 and 14q22-23), and RHPN2

(19q13). This overrepresentation of TGFβ pathway components suggests a key role for this pathway in CRC susceptibility (Tenesa and Dunlop, 2009). This is an attractive proposition given the critical role played by the pathway in mechanisms such as cellular differentiation, proliferation and migration (Alarmo and Kallioniemi, 2010). Recently, Carvajal-Carmona and colleagues (2011) partly confirmed the importance of the BMP and TGFβ pathway in CRC risk. In an analysis of ~25000 cases and an equal number of controls, two new CRC predisposition SNPs at *BMP4* (rs1957636) and *BMP2* (rs4813802) were identified that act independently to the original predisposition SNPs at these respective loci. In addition, the GWAS SNP rs4779584 at *GREM1* tags two further independent functional SNPs (rs16969681 and rs11632715). Taking into account the *SMAD* gene products that function within the BMP and TGFβ pathways (including SMAD7), the number of components within these pathways that influence risk includes 3 high-penetrance predisposition genes (*SMAD4, BMPR1A, GREM1*) and 7 low-penetrance variants at *BMP2, BMP4, GREM1, SMAD7* and *LAMA5*. A recent sequencing project, comparing the exomes of MSI and microsatellite stable (MSS) CRC tumours revealed that *BMPR1A* is mutated in both phenotypes highlighting a potential somatic role for this gene. Further functional work revealed that mutant protein is strongly impaired in its signalling function, and that stimulation with *BMP2* results in reduced maximum activity (Timmermann *et al.* 2010).

## 1.3 Colorectal tumourigenesis

The progression of CRC from adenoma to malignancy requires multiple mutations in distinct genes, the accumulation of which can take over a decade (Kinzler and Vogelstein, 1996). These genes have vital roles in controlling cellular processes such as the cell cycle and signalling pathways. Though different genes within the same pathway can be mutated in different colorectal cancers, they will generally be mutually exclusive within the same tumour (Vogelstein and Kinzler, 2004). An example of this are the *KRAS* and *BRAF* genes that are part of the ras-raf-MAP pathway downstream of the epidermal growth factor receptor (EGFR) (Crook *et al.* 2009). Though they are mutated in ~40-45% and ~10% of CRCs respectively, they generally do not co-occur. The stages of tumourigenesis are well defined and are highlighted in Figure 1.7.
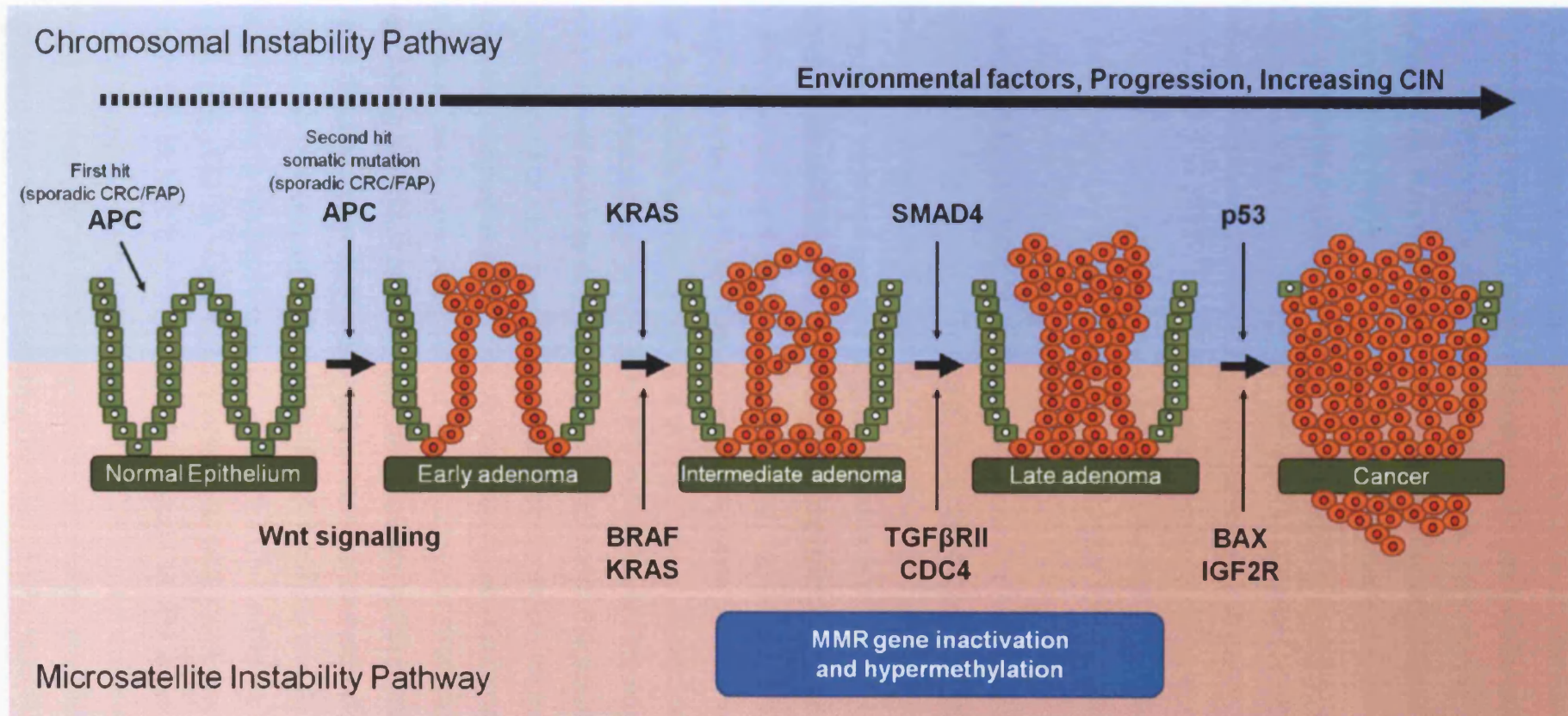
Figure 1.7 – Adenoma-carcinoma sequence depicting some of the key events which occur during colorectal tumourigenesis. Progression can occur via two main routes; the chromosomal and microsatellite instability pathways. The first step of tumourigenesis is the formation of an adenoma. Intermediate to late adenomas (and ultimately carcinomas) will generally acquire multiple hits in different pathways (e.g. the Ras-Raf-MAPK and/or TGFβ pathways). Some of these mutations occur often (e.g. KRAS is mutated in 40-45% of CRCs) while some might be less frequent but occur in a specific setting (e.g. mutations of BRAF are common in microsatellite unstable tumours).

## 1.3.1 Genomic Instability

The baseline mutation rates (as low as $10^{-9}$ per cellular generation (Albertini *et al.* 1990)) of the body are typically insufficient to account for the multiple mutations that are required in the development of cancer. Genomic instability is a characteristic of almost all human cancers and comes in various forms (Negrini *et al.* 2010) including chromosomal instability (CIN) and MSI. Abnormal chromosome numbers and structures, as well as abnormal mitoses are associated with CIN and result in genetic alterations, while MSI causes genetic alterations through defective DNA MMR (Söreide *et al.* 2009). Instability can also arise through epigenetic silencing of gene promoters. Through these mechanisms, genomic instability contributes to accumulation of mutations of genes involved in CRC (Grady and Markowitz, 2002).

The majority of CRCs (85%) develop according to the CIN pathway (Kinzler and Vogelstein, 1996) which can result from defects in a number of critical processes including chromosomal segregation, telomere stability and the DNA damage response, though there are likely to be other mechanisms that are as yet unknown. These defects ultimately result in gross changes in DNA by increasing the rate of chromosomal gains or losses (Kinzler and Vogelstein, 1996; Rajagopalan *et al.* 2003). Coupled with the karyotypic abnormalities typical in CIN tumours, is the non-random accumulation of a series of mutations in oncogenes and tumour suppressor genes. It is not clear which is the preceding event in tumourigenesis. It is possible that CIN creates the appropriate environment for the accumulation of such mutations but equally plausible that these mutations contribute to the de-stabilising of the genome. Regardless, the genetic changes appear to be an intrinsic feature of most cancers, as evidenced by recent high-throughput sequencing studies showing that mutations of DNA repair and replication genes are present in ~60% of cancer cell lines (Poulogiannis *et al.* 2010). Common events coupled with CIN include the allelic loss (loss of heterozygosity, LOH) of *p53*, *SMAD4*, and *APC* loci (Terdiman, 2000). Deletion of 1p and 8p is also common, though the genes involved with tumourigenesis at these loci are unknown. Along with the tumour suppressor genes described above, the *KRAS*, *CTNNB1*, and *PIK3CA* oncogenes at 12p12, 3p22 and 3p26 respectively are often mutated in CIN tumours (Santini *et al.* 2008; Sparks *et al.* 1998; Samuels *et al.* 2004). Several studies have revealed that CIN occurs at early stages of tumourigenesis (Bardi *et al.* 1997; Stoler *et al.* 1999), including that of

sporadic CRC (Shih *et al.* 2001) and hereditary CRC (Cardoso *et al.* 2006). There are a number of candidates for genes that, when mutated, contribute to the CIN phenotype. Indeed, more than 100 genes can lead to CIN in *Saccharomyces cerevisiae* (Pino and Chung, 2010) with only a minority of their human homologues having been implicated. Amongst the most likely candidates are genes in pathways that ensure correct mitotic segregation of chromosomes (including *BUB1, AURKA, PLK1, PTTG, APC*) and those involved in telomere regulation and DNA damage response (including *TERC, ATM, BRCA1/2, p53, MRE11*).

The next most common form of genomic instability is MSI. Genes that contain coding microsatellite repeats are prone to frameshift mutations (Söreide *et al.* 2006). For example, the *TGFβ-RII* gene is mutated in colorectal tumours with MSI with frameshift mutations occurring at specific hot-spots within the coding region (Markowitz *et al.* 1995). Of the 15% of CRCs that display MSI about 3% are associated with Lynch syndrome and 12% with sporadic CRC. A common feature of sporadic CRC is mutation of *BRAF* (Figure 1.7). A small proportion of CRCs display genomic instability that involves neither CIN nor MSI (Goel *et al.* 2003) but show CIMP.

## 1.3.2 Cancer genes

The genes that contribute to CRC fall into two separate classes – oncogenes and tumour suppressor genes (Vogelstein and Kinzler, 2002). An oncogene is a gene that when mutated becomes constitutively activated (or active when the wild type gene would not be). Mechanisms of activation of oncogenes include point mutation of residues critical in controlling the resulting proteins activity, chromosomal translocations (as caused by CIN), and gene amplifications (Vogelstein and Kinzler, 2004). Only one allele of an oncogene need be somatically mutated for activation. An example of an oncogene often mutated during the progression to colorectal tumourigenesis is *KRAS* (Figure 1.7). The *KRAS* gene encodes a protein critical in signal transduction through a number of intracellular pathways. As discussed in further detail later, the KRAS protein is a membrane-associated GTP-coupled protein that is activated by the binding of GTP enabling it to stimulate the MAPK cascade. The GTPase activity of KRAS then hydrolyses bound GTP to GDP, in-turn inactivating itself. KRAS acts as a signal switch molecule, and its mutation is

considered one of the main steps in the adenoma-carcinoma sequence of CRC genesis (Fearon and Vogelstein, 1990). Mutations in KRAS disable endogenous GTPase activity, in-turn leading to constitutive activation of downstream signalling and contributing to tumourigenesis (Forrester *et al.* 1987; Grady and Markowitz, 2002). This activation circumvents the need for upstream EGFR signalling. Other oncogenes that can contribute to CRC tumourigenesis include *CTNNB1* (Morin *et al.* 1997), *BRAF* (Davies *et al.* 2002), and *PIK3CA* (Samuels *et al.* 2006), with the *BRAF* protein having overlapping function with *KRAS*.

The other class of cancer genes are tumour suppressor genes. Unlike oncogenes, mutations of tumour suppressor genes lead to inactivation of the resulting protein – often because mutations lead to truncation or nonsense mediated decay (NMD) (Storey *et al.* 1998). Point mutations that target residues vital for the proteins function, along with insertions and deletions can result in loss of function. Other epigenetic mechanisms including hypermethylation (Baylin *et al.* 1998) and improper localization (Moll *et al.* 1992; Takahashi and Suzuki, 1994) can also contribute. Perhaps the most important tumour suppressor gene inactivated in CRC is *APC* since its loss represents one of the initiating steps of CRC (Powell *et al.* 1992; Groden *et al.* 1991; Kinzler *et al.* 1991). Another tumour suppressor gene, that is critical in the development of a number of cancers including colorectal, is *TP53* (Hollstein *et al.* 1991). Tumour suppressor genes act mostly in a recessive fashion – with mutations of both alleles being required for inactivation. In sporadic CRC, both of these mutations occur somatically, whilst in dominantly inherited forms of CRC one allele is mutated in the germline while the second is mutated somatically. In FAP, the *APC* gene has a germline mutation of one allele while a second somatic mutation of the other allele leads to CRC through adenoma formation (Fearnhead *et al.* 2001). This is in accordance with Knudson's two hit hypothesis (Knudson, 1996) (as portrayed in Figure 1.8).

A) Sporadic



Both alleles are normal → 1st sporadic hit on one allele → 2nd sporadic hit Inactivates second allele

B) Inherited

Inherited germline mutation

Every cell contains 1st germline hit → Both alleles are Inactivated by 2nd somatic hit → Loss of gene function → Progression to tumour formation
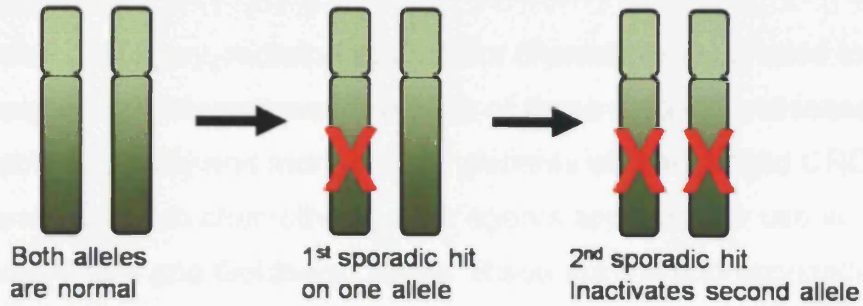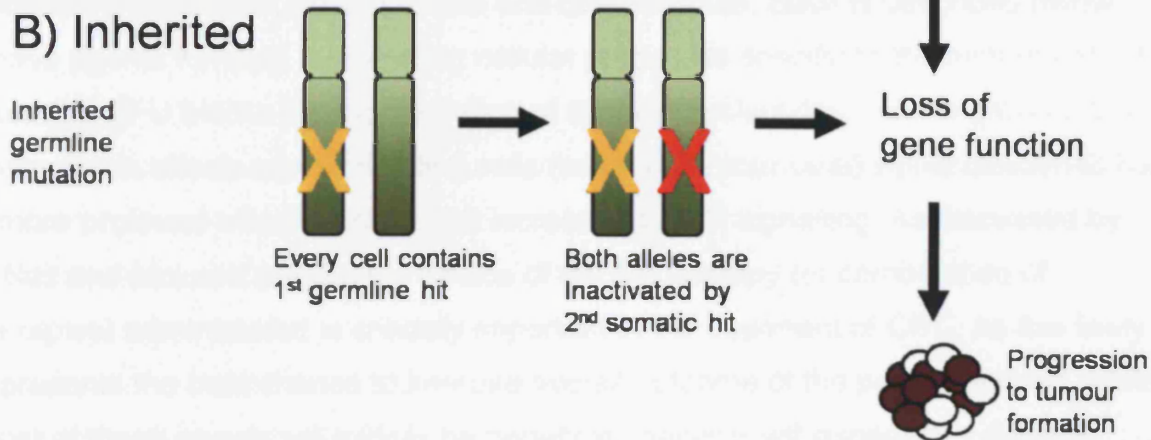
**Figure 1.8** – Diagram depicting Knudson's two hit hypothesis of tumourigenesis. In sporadic disease, two separate sporadic 'hits' are required – one per allele. In inherited disease, every cell will contain the first 'hit' so only one somatic hit is necessary for loss of gene function and progression to tumour formation.

## 1.4 Treatment of CRC

Currently, a combination of surgery, radiotherapy and/or chemotherapy is used in the treatment of CRC. Despite vast improvements in each of these areas, the disease still causes considerable morbidity and mortality. For patients with advanced CRC (aCRC), there are currently seven chemotherapeutic agents approved for use in routine clinical practice (O'Neil and Goldberg, 2005). These are the fluoropyrimidines 5-fluorouracil (5FU) and capecitabine, oxaliplatin, irinotecan and the monoclonal antibodies cetuximab, panitumumab and bevacizumab. Each is described below. These agents function by targeting cellular properties specific to the tumour cell – for example, 5FU blocks the incorporation of specific nucleotides into the growing DNA chain which affects rapidly dividing cells (such as tumour cells) whilst cetuximab has a more profound effect on cells with increased EGFR signalling. As discussed by O'Neil and McLeod (2006), the choice of the first therapy (or combination of therapies) administered is critically important in the treatment of CRC, as this likely represents the best chance to improve overall outcome of the patient. Indeed, whilst most of these agents will initially be beneficial, patients will experience disease relapse as drug-resistant tumour cells arise. Thus understanding of the agents most likely to elicit a response in patients, and the interactions between these agents, is of the utmost importance.

### 1.4.1 Standard chemotherapy

### 1.4.1.1 5-Fluorouracil

The synthesis of the fluorinated pyrimidine fluorouracil, by Heidelberger and colleagues more than 40 years ago, represents one of the most important landmarks in treatment of CRC (Heidelberger et al. 1957). This discovery was inspired by the observation that tumour cells in vitro showed greater uptake of uracil than normal cells. 5FU is inactive until it is metabolized in the cell, where it is converted into its active equivalent, fluorodeoxyuridine monophosphate (FdUMP), by the thymidine phosphorylase enzyme (TP). FdUMP, in the presence of the reduced folate 5,10-methylene tetrahydrofolate, forms a complex with, and inhibits the thymidylate synthase (TS) enzyme, in-turn inhibiting the synthesis of the thymine nucleotide that the tumour cell requires for DNA replication, and thus, cell division (Leichman, 2006). If not converted to FdUMP, 5FU is enzymatically degraded by dihydropyrimidine dehydrogenase (DPD) (McLeod et al. 1998; Omura, 2003).

Currently, 5FU is the therapeutic mainstay for CRC. Its development, along with that of other therapies used today has led to an increase in CRC median survival from ~6 months in untreated patients to ~20 months. 5FU is administered intravenously as it is variably absorbed by the gut and degrades rapidly. Folinic acid (FOL) (also known as Leucovorin) is used in addition to 5FU therapy and stabilises the ternary complex between 5FU and TS. Through clinical studies, it was found that the combination of 5FU with FOL doubles tumour response rates, though with little effect on OS (Doroshow et a, 1990; Advanced Colorectal Cancer Meta-Analysis Project, 1992).

### 1.4.1.2 Capecitabine

Capecitabine (Xeloda) is an orally administered alternative to bolus fluorouracil. It is rapidly taken up from the gut and metabolised, firstly through hydrolysis to 5'-deoxy-5-fluorocytidine (5'-DFCR) by hepatic carboxylesterase. This is then converted to 5'-deoxy-fluorouridine (5'-DFUR) by cytidine deaminase in the liver and the tumour cell, before undergoing a final transformation by TP at the tumour site to active 5FU (Schellens, 2007) which in-turn inhibits TS by the same mechanism described above. 5FU is preferentially generated at the tumour site due to the higher activity of the TP enzyme in tumour tissue compared to normal tissue (Miwa *et al.* 1998), in-turn minimising systemic exposure to 5FU. Numerous studies have shown it to be at least as active as bolus and infusional 5FU (Twelves *et al.* 2001; Van Cutsem *et al.* 2000; Cassidy *et al.* 2008), and suitable for use with oxaliplatin (Díaz-Rubio *et al.* 2002; Cassidy *et al.* 2004) and irinotecan (Patt *et al.* 2007; Bajetta *et al.* 2003; Borner *et al.* 2005; Rea *et al.* 2005). It has a favourable toxicity profile (Scheithauer *et al.* 2003; Van Cutsem *et al.* 2005) and is well received by patients who understandably prefer oral administration to i.v. treatments (Liu *et al.* 1997) which can have separate complications and inconveniences associated with central venous access (i.e. potential for venous thrombosis). Side-effects observed in capecitabine treatment include increased thromboembolic events, anaemia, diarrhoea, and hand-foot syndrome (HFS; Walko and Lindley, 2005). Hand-foot syndrome is a cutaneous condition that involves numbness of the hands and feet due to drug leakage in these extremities. At its least extreme, this condition displays discomfort that compromises daily life to a small extent, but in worst cases the skin displays blistering, ulceration and/or pain that can severely affect day-to-day activities. For this reason, dosage adjustment is recommended.

In the last decade, two new cytotoxic agents have emerged in the treatment of CRC;

### 1.4.1.3 Oxaliplatin

Oxaliplatin is a diaminocyclohexane-platinum analogue that has been extensively studied through Europe and the USA and found to have significant activity in CRC (Bleiberg 1998). When administered, oxaliplatin interacts with plasma proteins and is distributed throughout the tissues of the body. About 50% of the drug is excreted from the body via the kidneys (Comella *et al.* 2009). It works by alkylating DNA, forming adducts that inhibit replication and transcription (Woynarowski *et al.* 1998). More specifically, these adducts comprise cross-links between the activated platinum species and specific base sequences – usually two adjacent guanine bases or adjacent guanine and adenine bases (Fink *et al.* 1997). Oxaliplatin activates numerous pathways including the p38 and PI3K-PTEN-Akt pathways (Rakitina *et al.* 2003), as well as the caspases cascade (Griffiths *et al.* 2004), while damage that results from Oxaliplatin is repaired by members of the nucleotide excision repair pathway (Arnould *et al.* 2003).

### 1.4.1.4 Irinotecan

Irinotecan (also known as CPT-11) is a semi-synthetic analogue of camptothecin that works by inhibiting the nuclear enzyme topoisomerase I (topo I) causing S-phase-specific cytotoxicity (Liu *et al.* 2000). Irinotecan must undergo hydrolysis or de-esterification to form its active metabolite SN-38 which inhibits topoI to an even greater extent (Rothenberg, 1997). Carboxylesterase is required for this step. TopoI normally functions in relieving torsional stress in front of the replication fork, as DNA unwinds for normal replication, transcription and repair recombination. To do this, TopoI cleaves the DNA backbone and forms a covalent enzyme-DNA linkage, facilitating the passing of the intact strand through the break to relieve this tortional stress, before finally re-sealing the break. These stabilised single-strand breaks are thus reversible and do not threaten the cell. However, should an advancing replication fork encounter a single-strand break, a double strand break will form. These induce phosphorylation of the H2AX histone variant, activate various signalling pathways, lead to the irreversible arrest of the replication fork, and cell death (Liu *et al.* 2000; Paillas *et al.* 2010). Thus, inhibition of TopoI by Irinotecan prevents ligation of the nicked strands that are met by the advancing replication fork that in-turn form irreversible double strand breaks. Evidence also suggests that,

when present at high concentrations, Irinotecan can induce apoptosis independent of the S-phase through a transcriptionally mediated DNA damage model (Morris and Geller, 1996)

## 1.4.2 Targeted therapies against CRC

In CRC treatment, there are two major strategies of targeted therapies - those that target angiogenesis and those that inhibit the EGFR pathway. EGFR is a member of the erbB or HER transmembrane protein kinase receptor family that acts as the gateway for multiple downstream, intracellular signalling pathways including the Ras-Raf-MAP and PI3K-PTEN-Akt pathways. Through these pathways, EGFR regulates multiple processes including apoptosis, cellular growth, proliferation, differentiation, and migration (Woodburn *et al.* 1999; Scaltriti *et al.* 2006). Several EGFR ligands have been identified including EGF itself, heparin-binding EGF-like growth factor, transforming growth factor (TGF), epiregulin, betacellulin, and amphiregulin. The receptor is anchored in the cytoplasmic membrane and has a structure consisting of an extracellular ligand-binding domain, a short hydrophobic transmembrane region, and an intracellular tyrosine kinase domain (Yarden *et al.* 2001; Hynes *et al.* 2005). Binding of ligand to the receptor induces a conformational change of the receptor ectodomain, allowing receptor dimerisation and autophosphorylation of several tyrosine residues within the COOH (Carboxy)-terminal of the receptors (Burgess *et al.* 2003; Hubbard *et al.* 2005). This phosphorylation provides docking sites for cytoplasmic proteins of the Ras-Raf-MAP and PI3K-PTEN-Akt pathways. EGFR is up-regulated in 60-80% of CRCs (Untawale *et al.* 1993; Salomon *et al.* 1995), correlating with reduced survival (Mayer *et al.* 1993). Given this, and its central role in normal, and tumorigenic growth, EGFR has been identified as a valid target for multiple cancers. In CRC treatment, the monoclonal antibodies cetuximab (Erbitux[®,] Merck KGaA) and panitumumab (Vectibix[®], Amgen) have been studied as agents that blockade EGFR signalling.

## 1.4.2.1 Cetuximab

Cetuximab is a recombinant chimeric monoclonal IgG1 antibody that binds the extracellular domain of human EGFR, in doing so preventing ligands EGF and TGF-α from binding the receptor, and triggering receptor internalisation (Sunada *et al.* 1986); thus inhibiting downstream signalling. This is summarised in Figure 1.9. The mechanisms by which cetuximab brings about tumour cell death include blocking of cell cycle progression (Wu *et al.* 1996; Peng *et al.* 1996; Kiyota *et al.* 2002), promotion of apoptosis (Wu *et al.* 1995; Mandal *et al.* 1998; Liu *et al.* 2000; Sclabas *et al.* 2003), inhibition of angiogenesis (Petit *et al.* 1997; Perrotte *et al.* 1999; Bruns *et al.* 2000; Martinelli *et al.* 2010; Pueyo *et al.* 2010) and metastasis (Huang *et al.* 2002; O-charoenrat *et al.* 1999; O-charoenrat *et al.* 2000; Bruns *et al.* 2000) and indirectly through antibody mediated cell cytotoxicity (Kawaguchi *et al.* 2007).

## 1.4.2.1.1 Side effects of cetuximab

Given the ubiquitous presence of EGF receptors in many tissues, the potential for a variety of cetuximab related adverse reactions is high. The most common side-effect associated with administration of cetuximab involves dermatological reactions – Papulopustulous rash and paronychia present within days of treatment, followed by atrophy of the skin (Arnold and Seufferlein, 2010). An estimated 70-100% of patients will present with such reactions (Fox, 2006) which reach maximum intensity after 2 to 3 weeks (Grothey, 2006). This is not surprising considering that the skin contains an abundance of EGF receptors – maintaining the integrity of the epidermis. Locations of these rashes most commonly include the face, scalp, and upper torso. Other side-effects have also been observed, including alopecia (Martin-Martorell *et al.* 2008), eyelash trichomegaly (Cohen *et al.* 2011), telangiectasias, ocular toxicities, hypomagnesemia, diarrhoea (Fakih and Vincent, 2010) and severe hypersensitivity reactions (1-22% of patients; Chung *et al.* 2008)
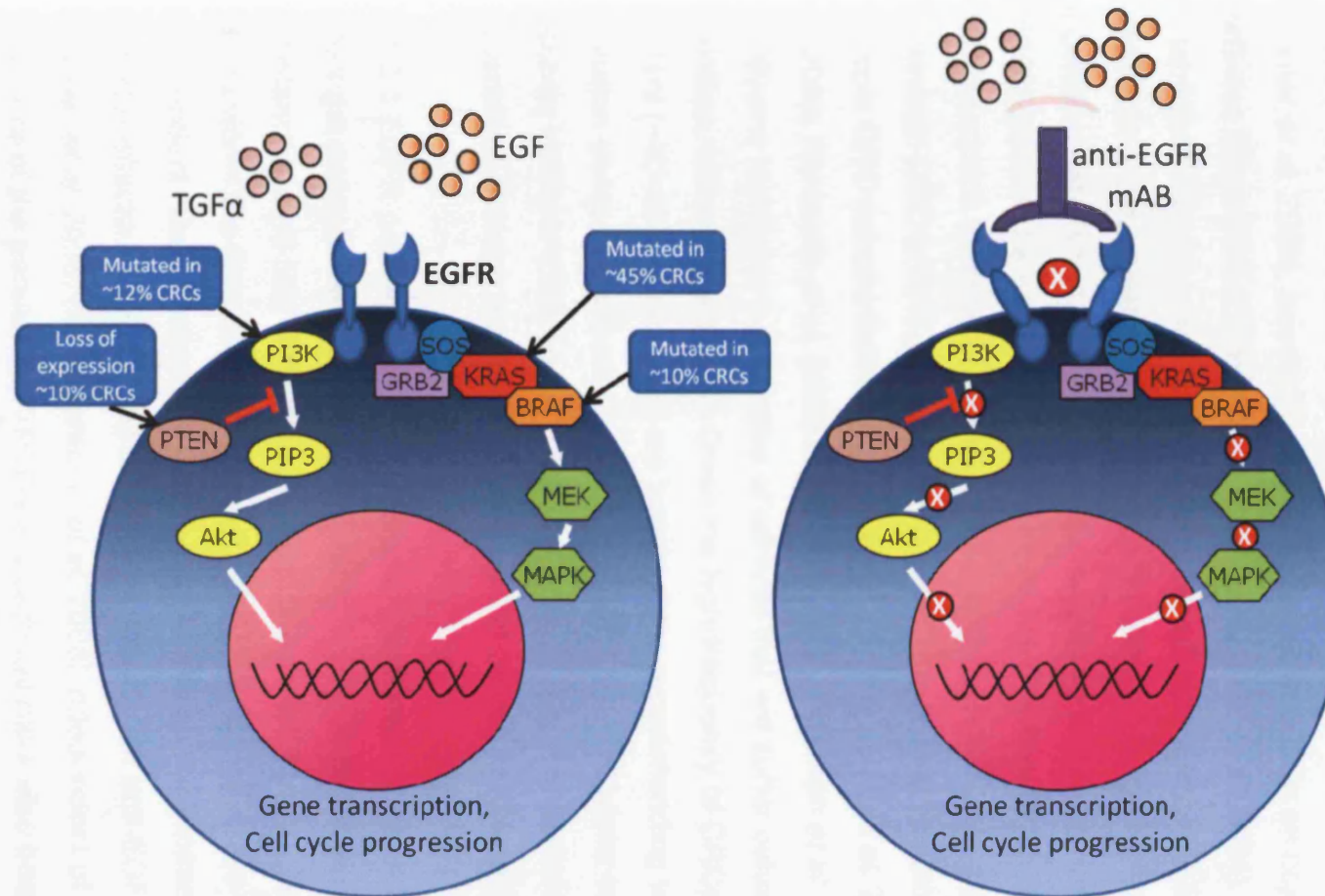
**Figure 1.9** – Diagrammatic representation of the mechanism of action of cetuximab. Normally EGFR is bound by its ligands such as TGFα and EGF, triggering a cellular signalling cascade that contributes to the regulation of gene transcription and cell cycle progression. Cetuximab is a monoclonal antibody that competitively binds to the extracellular component of the EGFR, blocking ligand bindings. This in-turn blocks the signalling cascade and as such gene transcription and cell cycle progression are blocked within the tumour cell, killing it.

## 1.4.2.2 *KRAS* and response to anti-EGFR therapies

Despite encouraging initial findings that cetuximab could be used successfully as a monotherapy (Jonker *et al.* 2007; Saltz *et al.* 2004) or in combination with irinotecan (Cunningham *et al.* 2004; Wilke *et al.* 2008; Sobrero *et al.* 2008; Arnold and Seufferlein, 2010), oxaliplatin/5FU (Arnold *et al.* 2008) or oxaliplatin/capecitabine (Borner *et al.* 2006), benefit from the drug is still limited to an objective response rate between 9 and 20% (Cunningham *et al.* 2004; Saltz *et al.* 2004). It was soon established that response to cetuximab was influenced by *KRAS* somatic mutation status. As such, mutations of *KRAS* can be referred to as predictive markers; that is a genetic feature that alters a patients response to treatment. Lievre et al (2006) initially observed a significant difference in overall survival (OS) between patients wild-type (16.3 months) and mutant (6.9 months) for *KRAS* in a group of 30 metastatic CRC (mCRC) patients (P=0.016). This finding has since been validated in multiple CRC cohorts that tested cetuximab (Van Cutsem *et al.* 2008; De Roock *et al.* 2008; Karapetis *et al.* 2008) or panitumumab (Freeman *et al.* 2008; Amado *et al.* 2008) and highlights a proportion of patients that will suffer cetuximab associated toxicities, with no real benefit. Given the high-frequency of CRCs that are *KRAS* mutant (~40-45%), guidelines are now in place recommending testing of *KRAS* mutation status (specifically 'hot-spot' codons 12 and 13) prior to treatment of mCRC patients with anti-EGFR agents (Allegra *et al.* 2009). *KRAS* mutation status can be ascertained through analysis of formalin-fixed paraffin-embedded (FFPE) tissues.

## 1.4.2.3 EGFR pathway components and resistance to cetuximab

Amongst patients wild-type for *KRAS*, up to 65% (Allegra *et al.* 2009) still show resistance to anti-EGFR therapies, raising the possibility that the mutation of other genes within the Ras-Raf-MAPK and PI3K-PTEN-Akt pathways could play an independent role in resistance. *BRAF*, a serine-threonine kinase that acts as the principle effector of *KRAS*, has also been implicated in anti-EGFR therapy resistance (Jhawer *et al.* 2008; Di Nicolantonio *et al.* 2008) independent of *KRAS* status. Members of the parallel PI3K-PTEN-Akt pathway have also been studied as candidates of resistance. Studies have shown contrasting results for *PIK3CA*; Perrone et al (2009) and Jhawer et al (2008) have offered evidence for resistance in the *PIK3CA* mutant setting, while Prenen et al (2009) failed to show an effect. Also, loss of *PTEN*, a lipid phosphatase that regulates the PI3K-Akt pathway, may lead to

aberrant Akt signalling, and thus resistance to cetuximab (Frattini *et al.* 2007; Perrone *et al.* 2009; Razis *et al.* 2008; Bouali *et al.* 2009; Loupakis *et al.* 2009; Sartore-Bianchi *et al.* 2009; Laurent-Puig *et al.* 2009; Negri *et al.* 2010; Kim *et al.* 2010). Cell lines that are both Ras/*BRAF* mutant and *PIK3CA* mutant/PTEN null show maximal resistance to cetuximab as compared to those with mutations in just one of the two parallel pathways (Jhawer *et al.* 2008).

*NRAS* and KRAS are two of three different isoforms of the Ras protein, with *HRAS* being the third. Despite their structural similarity, the Ras isoforms have been shown to differ in their interactions with effectors (Yan *et al.* 1998) and in their importance for viability; in mice knock-out experiments, *KRAS* is essential for development but *HRAS* and *NRAS* are not (Johnson *et al.* 1997). Whilst *NRAS* is mutated far less frequently in CRC, it may still play a role in altering response to cetuximab; patients with *KRAS, BRAF* and *NRAS* wild-type tumours show significant association with outcome after cetuximab treatment (Lambrechts *et al.* 2009). *HRAS* mutations are not present in CRC (Bamford *et al.* 2004) and thus do not play a part in anti-EGFR therapy resistance.

Recently, De Roock et al (2010a) carried out a thorough combined analysis of the *KRAS, BRAF, NRAS* and *PIK3CA* genes, and their role in altering response of chemotherapy refractory metastatic CRC patients, to cetuximab plus chemotherapy. The primary analysis of this study was objective response, as opposed to the usual OS, which can be affected by prognostic markers, and progression free survival (PFS) (the nature of the study did not allow this). Within *KRAS* mutant samples, the response rate was 6.7% compared with 35.8% in *KRAS* wild-type samples (OR=0.13, 95% CI 0.07-0.22, P<0.0001). Within this *KRAS* wild-type group, patients mutant for *BRAF* had a response rate of 8.3% that was significantly different when compared to *BRAF* wild-type patients (38%) (OR=0.15, 0.02-0.51, P=0.0012). This was the same for *NRAS* where mutant patients had a response rate of 7.7% vs 38.1% in wild-types (OR=0.14, 0.007-0.70, P=0.013). Results for *PIK3CA* proved to be particularly interesting; mutations within exon 9 of the gene were found to have no effect on cetuximab, but mutations of exon 20 did play a role in worsening response to cetuximab (0% vs 36.8%, P=0.029). OS and PFS were analysed as secondary analyses, and results were similar in each case. Based on these results the authors

recommend the ordered testing of *KRAS, BRAF, NRAS* and *PIK3CA* exon 20 for somatic mutations within patients, suggesting an objective response rate of 41.2% in this 'all wild-type' group. Whilst interesting, the staggering differences in performance between mutant and wild-type groups suggests that the findings may still be influenced by the strong prognostic value these mutations carry, even though the authors tried to stratify for this effect.

Given that EGFR is the target of this antibody and that it is over-expressed in such a high proportion of CRCs, most early studies of cetuximab efficacy required confirmed *EGFR* expression. Moroni and colleagues (2005) observed a significant difference (P<0.0001) in increased *EGFR* copy number status, as analysed by fluorescence *in situ* hybridisation (FISH), between responders and non-responders. This result suggested that patients could be selected based upon their *EGFR* copy number. However Personeni and colleagues (2008) warned that, whilst *EGFR* copy number is correlated with response independent of *KRAS* status (p=0.016) and OS (p=0.005), the published criteria used in determining copy number status might lead to mis-classification – for this reason clinical decision making based upon this factor is not warranted. Furthermore, studies have attempted but failed to validate the correlation between *EGFR* expression and response to cetuximab (Cunningham *et al.* 2004; Folprecht *et al.* 2010).

### 1.4.2.4 Alternative factors affecting response to EGFR targeted treatments

A further predictor of response to cetuximab is its associated side-effect involving skin rashes. Saltz et al (2003 and 2004) were amongst the first to suggest that the rash might be a measure of sensitivity, after observing superior survival in patients with a rash as compared to those without. This observation was also confirmed by Cunningham and colleagues (2004). A possible explanation for this phenomenon may lie in the presence, or lack thereof, of a CA dinucleotide repeat within the first intron of the *EGFR* gene. Low numbers of this repeat are associated with this rash. Studies of head-and-neck cancer cells with a lower number of the dinucleotide had high expression levels of *EGFR* and were more sensitive to another inhibitor of EGFR activity, erlotinib (Amador *et al.* 2004). Steps can be taken to lower the rate of severe skin toxicity without compromising the therapeutic efficacy of anti-EGFR therapies (Arnold and Seufferlein, 2010). As with EGFR copy number status, a lack

34

of uniformity in grading, combined with the fact that time to skin manifestation is so variable, means that rashes are not used in clinical decision making.

### 1.4.2.5 Panitumumab

Panitumumab is a human IgG2 monoclonal antibody that, like cetuximab, is directed against EGFR. Panitumumab blocks binding of the EGF and TGF-α ligands, and leads to internalisation of the receptor. Unlike cetuximab, panitumumab induces arrest of the cell cycle during the G0-G1 interphase. Also, it does not induce the antibody-dependent cell-mediated cytotoxicity (ADCC) mechanism (Berardi *et al.* 2010). It has been the subject of a number of pre-clinical and clinical studies that have suggested it to have efficacy as a monotherapy or in combination, in the treatment of patients with mCRC (Van Cutsem *et al.* 2007; Van Cutsem *et al.* 2008; Hecht *et al.* 2009; Siena *et al.* 2009; Peeters *et al.* 2009), in a *KRAS* dependent manner (Amado *et al.* 2008; Douillard *et al.* 2009; Peeters *et al.* 2009). As with cetuximab, there is a correlation between skin toxicity and response to panitumumab (Van Cutsem *et al.* 2007; Peeters *et al.* 2009; Berlin *et al.* 2007).

Many oncologists view cetuximab and panitumumab as being interchangeable though a lack of phase II/III data on combined panitumumab therapy means that it is usually administered in isolation (Kim, 2009). Studies have attempted to determine if panitumumab can be administered as a monotherapy after cetuximab failure. Metges and colleagues (2010) show that there is promise for this approach, with 72.7% of trial patients (albeit 32 patients in total) showing clinical benefit. Other small studies have largely agreed with this observation (Saif *et al.* 2010; Brugger, 2010) but further work is needed to sufficiently validate this is a suitable treatment option.

### 1.4.2.6 Bevacizumab

Bevacizumab is a humanised monoclonal antibody that targets the vascular endothelial growth factor-A (VEGF-A), thus inhibiting blood vessel formation. Though it has little efficacy as a single agent, it has displayed promise when used in combination with other standard chemotherapeutic agents used in the treatment of mCRC (Arnold and Seufferlein, 2010). The first phase II trial to show this involved the comparison of 5FU/FOL alone with 5FU/FOL plus bevacizumab in 104 mCRC patients (Kabbinavar *et al.* 2003). It showed improved RR, PFS and OS in the

combination arm and was followed by a series of other first- and second-line therapy studies (Hurwitz *et al.* 2004; Kabbinavar *et al.* 2005; Giantonio *et al.* 2007; Saltz *et al.* 2008) that showed efficacy with a range of alternative partners including oxaliplatin and irinotecan.

Whilst the inhibition of individual signalling pathways has shown clear beneficial effects, the question remains as to whether multiple signalling pathways can be targeted at once. This seems a reasonable hypothesis based on the fact that tumours often have multiple genetic changes that affect multiple pathways. Studies have trialled the combined use of cetuximab/panitumumab, bevacizumab and standard chemotherapy in the treatment of mCRC (Ciardiello *et al.* 2000; Jung *et al.* 2002; Shaheen *et al.* 2001; Tonra *et al.* 2006; Saltz *et al.* 2007). Despite encouraging initial findings little evidence has been shown for improved survival in this combined treatment approach (Tassinari *et al.* 2010) regardless of *KRAS* status (Tol *et al.* 2009).

## 1.5 Prognostic factors for CRC

Prognostic markers provide information concerning a patient's outcome regardless of treatment given. Individually and collectively, prognostic factors have the potential to guide clinical decision making for administration of adjuvant therapy – especially within patients of stage II CRC where the decision to treat or not treat can still dramatically alter disease course. Of note, biomarkers can be both prognostic and predictive (e.g. hormone receptor status in breast cancer) (Oldenhius *et al.* 2008). The only routinely used prognostic marker for CRC is clinical stage, which combines depth of tumour invasion, nodal status and distant metastasis (Walther *et al.* 2009). Other factors thought to influence prognosis include lifestyle (Haydon *et al.* 2006; Reeves *et al.* 2007), systemic inflammatory response to the tumour (Leitch *et al.* 2007) and the tumour immunologic microenvironment (Galon *et al.* 2006).

Protein and genetic markers have been studied in an attempt to refine prognostic prediction for CRC. However, none have been validated sufficiently for clinical practice. Some progress has been made in other cancers though – for example,

hormone receptor status has been associated with prognosis in breast cancer (Dunnwald et al. 2007).

## 1.5.1 Somatic prognostic markers

Somatic genetic changes are perhaps the most studied markers for prognostic influence (Table 1.2) since they offer the most biological rationale for a role in defining the course of a cancer. Whilst somatic effects have been found for CRC, the results of different studies of specific markers have often been contradictory. The prognosis of patients with MSS CRC is stage and grade dependent and tumours with identical morphological features display considerable heterogeneity in terms of clinical outcome. Meanwhile, a meta-analysis of over 7000 CRC patients, found that patients with MSI-positive tumours had significantly improved OS as compared to those with MSS CRCs (Popat et al. 2005). MSI could be used as a prognostic marker since there are reliable assays for determining MSI status (Pritchard and Grady, 2011).

Further studies of prognostic markers have focused on events that occur during the adenoma to carcinoma progression of CRC, as presented in Figure 1.7. It has been suggested that APC might be suitable as a prognostic marker – a study of 218 CRC patients stratified for MSI status suggested that mutations abolishing β-catenin binding sites of the APC gene correlated with shorter survival as compared to mutations elsewhere in the gene (P=0.045; Løvig et al. 2002). KRAS mutations also play a key role in colorectal tumourigenesis and any prognostic role for it would be particularly advantageous since it is already routinely analysed as a marker of cetuximab efficacy. A large international consortium has reported that somatic KRAS mutations confer poor prognosis (Andreyev et al. 1998), although subsequent analyses have suggested that this outcome is limited to the G12V substitution and to Stage III patients (Andreyev et al. 2001). Other studies have found no such associations (Ince et al. 2005; Ogino et al. 2009; Samowitz et al. 2000; Wang et al. 2003). BRAF mutations (specifically the common V600E mutation) has been reported by various groups to influence survival in patients with different stages of disease (Ogino et al. 2009; Farina-Sarasqueta et al. 2010). In addition, PIK3CA mutations have been found to influence patient survival (Kato et al. 2007; Ogino et al. 2009).

## Somatic prognostic markers

| Biomarker | Mutation Frequency (%) | Prognosis | Evidence | Status | Reference |
|-----------|------------------------|-----------|----------|--------|-----------|
| MSI | ~15 | Favourable | Strong | Routine assays available but not yet used | Popat et al. 2005 |
| CIN | ~85 | Unfavourable | Strong | No reliable/robust assays available | Walther et al. 2008 |
| APC (mutations of β-catenin binding regions) | >90% | Unfavourable | Limited | Not used | Løvig et al. 2002 |
| KRAS | 40-45 | Unfavourable | Moderate | Not used | Andreyev et al. 1998 |
| BRAF | ~10 | Unfavourable | Moderate/Strong | Not used | Ogino et al. 2009 |
| PIK3CA | ~20 | Unfavourable | Limited | Not used | Ogino et al. 2009 |
| TP53 | ~40 | Unfavourable | Limited | Not used | Russo et al. 2005 |
| 18q LOH | 50 | Unfavourable | Moderate | Not used | Roth et al. 2009 |

**Table 1.2** – Summary of somatic markers. Shown are select examples of genes that have been suggested to have a potential prognostic influence.

Other common events that occur during CRC progression, such as mutations of *TP53* (Russo *et al.* 2005; Munro *et al.* 2005) and loss of 18q (Popat and Houlston, 2005; Halling *et al.* 1999; Roth *et al.* 2009) have been studied for prognostic potential with mixed success. In the case of 18q LOH, any prognostic effect observed may simply represent the molecular phenotype of that particular tumour, since its loss correlates so tightly with CIN (Walther *et al.* 2009).

## 1.5.2 Germline prognostic markers

The search for germline prognostic factors has primarily focussed on the pharmacological pathways involved in the metabolism and mechanism of action of the chemotherapeutic agents used in the treatment of CRC (Table 1.3). For example, inherited variants in the TS gene, particularly for patients receiving adjuvant therapy with 5FU, have been suggested to act as prognostic factors (Dotor *et al.* 2006; Marcuello *et al.* 2004). The Glutathione S-transferase (GST) genes are involved in the detoxification of chemotherapeutic agents. A study of 338 CRC patients suggested that copy number variation of GSTM1 associated with survival in patients that received chemotherapy (Funke *et al.* 2010). More recently, investigators have sought prognostic factors in pathways that are likely to influence tumour progression. For example, the *VEGF* superfamily of endothelial growth factors has been shown to critically influence tumour-related angiogenesis and *VEGF* polymorphisms have been associated with survival outcome (Kim *et al.* 2008).

Common polymorphisms of repair pathways genes have been studied for effects on patient survival. A study investigating the prognostic impact of 68 SNPs from 7 MMR genes in 2,068 CRC patients found that the *MSH3* intronic SNP rs863221 altered patient survival (Koessler *et al.* 2009). Similarly, a screen of 377 CRC cases was also carried out for prognostic factors from BER, nucleotide excision repair (NER) and double-strand break repair (DSBR) pathways (Moreno *et al.* 2006). It showed that three polymorphisms were associated with better prognosis (*XRCC1* R399Q, *XRCC3* T141M and *MGMT* L84F) and one with worse (*ERCC1* 19007T>C) though only the first and last SNPs remained significant after adjustment for multiple testing. ERCC1 is an excision nuclease within the NER pathway that has been associated with survival in patients administered oxaliplatin (Shirota *et al.* 2001) and 5FU (Metzger *et al.* 1998) based therapies. The ERCC1-118 C/T (rs11615) polymorphism

has also been associated with reduced OS in mCRC patients by some groups (Chua *et al.* 2009; Ruzzo *et al.* 2007; Stoehlmacher *et al.* 2004), but not by others (Pare *et al.* 2008; Viguier *et al.* 2005).

The above studies of prognostic markers for CRC have all been hypothesis driven – that is, targeted against genes already known to play some part in CRC progression and/or drug metabolism. The findings from these studies could be truly transformative for a number of reasons. Firstly, whilst the search for prognostic factors has largely focused on molecular markers in tumour tissue, germline markers have potential to overcome a number of issues facing the use of somatic prognostic predictors (e.g. germline DNA is easier to obtain than tumour DNA). Secondly, the identification of survival biomarkers might provide further insight into the mechanisms through which cancers progress and metastasise. Thirdly, the biomarkers identified (and the mechanisms through which they exert their effects) might represent novel targets for therapeutic intervention. Finally, survival biomarkers will help to inform clinical decision making and ultimately decrease mortality associated with cancer.

## Germline prognostic markers

| Biomarker | Variant | Risk allele Frequency (%) | Prognosis | Reference |
|---|---|---|---|---|
| *TS* (1494del6) | 1494del6 | ~36* | Favourable | Dotor *et al.* 2006 |
| *GSTM1* (copy number) | Copy number variation | Unknown | Favourable | Funke *et al.* 2010 |
| *VEGF* | -634G>C | 20-40 | Favourable | Kim *et al.* 2008 |
| | +936C>T | ~20 | Unfavourable | |
| *MSH3* | rs863221 | ~35 | Favourable | Koessler *et al.* 2009 |
| *XRCC1* | R399Q | ~40-45 | Favourable | Moreno *et al.* 2006 |
| *XRCC3* | T141M | ~61 | Favourable | Moreno *et al.* 2006 |
| *MGMT* | L84F | ~4 | Favourable | Moreno *et al.* 2006 |
| *ERCC1* | 19007T>C | 60 | Unfavourable | Moreno *et al.* 2006 |
| | rs11615 | ~35 | Unfavourable | Chua *et al.* 2009 |

**Table 1.3** – Summary of germline variants. Shown are select examples of genes (and variants) shown to have a potential prognostic influence.

## 1.6 Somatic mutation detection

The ability of *KRAS* mutations to predict lack of response to cetuximab has led to the recommendation of somatic mutation testing by the European Medicines Agency (Becquemont *et al.* 2011). Prior to this, *KRAS* had been screened as part of experimental and clinical trials but not in routine testing. Now somatic analysis of *KRAS* is wide spread in clinical laboratories. However, before mutation analysis can be carried out, tumour material must be obtained.

In brief, tumour material is removed via surgery (resection or biopsy) and embedded in paraffin as FFPE blocks. Sections can be cut from these blocks, stained and visualised for tumour identification. Tumour material is removed from matched unstained slides and DNA is extracted. An important consideration when carrying this out is the effect of interfering normal tissue. If DNA from normal tissue is present in the test sample, it will mask the presence of mutant allele. Thus it is important to capture as pure a population of tumour cells as possible. A variety of techniques are employed in the removal of tumour material from paraffin sections, including macrodissection and laser capture microdissection (LCM). Even using these techniques, there will often be some level of contaminating normal material. Also when analysing oncogenes, each tumour cell will likely still carry one wild-type allele. As such, mutation analysis must still be sensitive enough to detect mutant allele against a backdrop of normal allele. Examples of sections abundant in tumour material and with sparse pockets of tumour material are shown in Figures 1.10A and 1.10B. Once tumour material has been obtained, DNA must be extracted for mutation analysis.
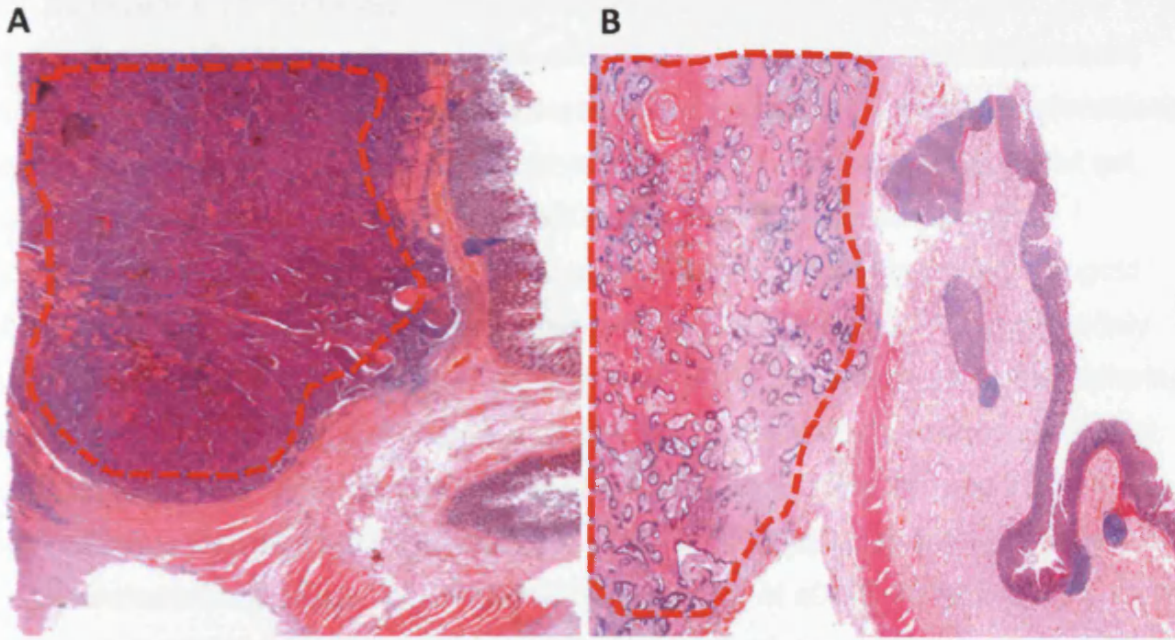
**Figure 1.10** – Haematoxylin and Eosin (H&E) stained colon sections containing varying amounts of tumour material (areas containing tumour are highlighted with a red, dotted area). (A) Here tumour tissue is abundant and can be easily extracted without collecting too much interfering normal tissue. The most appropriate method of tumour extraction here would be macrodissection. (B) Here tumour material is found in small and sparse globules that are not easy to extract without also picking up normal tissue. In order to ensure accurate collection of tumour cells, a precise cutting technique such as LCM would be most appropriate.

### 1.6.1 Current mutation detection technologies

A variety of methods have been employed for general mutation analysis in recent years. Though conformation-based separation techniques such as high performance liquid chromatography (HPLC; Mascarenhas et al. 2009), denaturing gradient gel electrophoresis (DGGE; Hayes et al. 2000) and single strand conformation polymorphism (SSCP; Dillon et al. 2001) are used in mutation screening, the gold standard for mutation analysis remains sequencing technologies. The most widely used of these is dideoxy sequencing (Brink et al. 2003). However, as well as offering limitations in throughput and cost, this technology is not sensitive enough to detect mutant allele in samples containing little tumour material. Thus, the use of technologies that detect low levels of mutant allele has exploded in recent years. These include Solid phase minisequencing (Ihalainen et al. 1994; Paunio et al. 1996), digital PCR (Vogelstein and Kinzler, 1999), melting curve analysis (Pritchard et al. 2010), and next-generation sequencing (NGS) (Peeters et al. 2010). Each has inherent advantages and disadvantages, with sensitivities, labour requirements, turn-around time, cost and potential for automation and multiplexing (Arcila et al. 2011).

Two alternative techniques are used in this thesis and as such as described in greater detail below;

### 1.6.1.1 Pyrosequencing

Pyrosequencing is a real-time, non-electrophoretic technique that relies upon the generation of pyrophosphate as nucleotides anneal to a growing template DNA strand. Nucleotides are dispensed in a user-defined order into the reaction mix, with those that represent the binding partner of the next nucleotide in the template strand annealing. Nucleotides that do not anneal (i.e. not complimentary or excess nucleotide) are degraded by apyrase. As nucleotides anneal to a growing synthesis strand, pyrophosphate is generated. This is detected as a fluorescent emission, the intensity of which gives an indication of the amount of annealed nucleotide. The overall process of Pyrosequencing is highlighted in Figure 1.11.

There are a number of advantages in the use of Pyrosequencing. Firstly it offers a simple, rapid and quantitative detection system for known mutations and SNPs (Poehlmann et al. 2007; Fakhrai-Rad et al. 2002). It can also be used in the

44

quantitative analysis of CpG island methylation events (Colella *et al.* 2003; Tost *et al.* 2003). Unlike other sequencing based methods, Pyrosequencing does not require the usual clean up step. Instead it requires that the initial PCR be carried out with one biotinylated primer. The subsequent biotinylated amplification product is immobilised on streptavidin-coated sepharose beads that are captured onto probes using Qiagen's vacuum prep tool, and transferred into the reaction mix for sequencing. Also, since the sequencing primer can be placed adjacent to the sequence of interest, the overall fragment size of the PCR product does not have to be that big, which is desirable when dealing with fragmented FFPE DNA (Okayama *et al.* 2010). Though the typical read length of Pyrosequencing is shorter than that of traditional dideoxy sequencing (40-50 nucleotides), this is not an issue when screening *KRAS* since mutations cluster at hot-spots. The technology allows for the sensitive analysis of multiple samples (utilising a 24 or 96 well format) in a short period which is desirable for clinical use (Poehlmann *et al.* 2007). In recent years, it has been used by various groups in the analysis of *KRAS* (Ogino *et al.* 2005; Poehlmann *et al.* 2007; Weichert *et al.* 2010; Tsiatis *et al.* 2010) with mutation detection sensitivities as low as 3-5% mutant allele.

## 1.6.1.2 Sequenom

The Sequenom platform is a chip-based matrix-assisted laser desorption-time-of-flight (MALDI-TOF) mass spectrometric SNP genotyping technology that relies upon single base primer extension, with A, C, T, or G terminators, through the mutation site. This generates allele-specific extension products that are spotted onto spectro-chips for analysis (Figure 1.12). The difference in mass between mutant and wild-type product is resolved by MALDI-TOF, allowing the quantitative comparison of alleles.

Originally designed for use in SNP genotyping of amplified DNA fragments (Nelson *et al.* 2004), it has become established in the quantitative analysis of known somatic mutations as well as methylation events (Izzi *et al.* 2010). It has a high multiplexing capacity that can be critical when dealing with small amounts of precious clinical samples. This is further helped by the fact that only small amounts of input DNA (from FFPE tissues) are required (5ng). Assays are limited to 40 reactions due to the fact that the mass spectrophotometer has a limited mass detection range which can

only facilitate a certain number of reaction product sizes without overlap occurring. Complications such as the increased chance of primer interaction and decreased amplification efficiency also impact on the number of reactions that can be run. Regardless, the ability to multiplex allows for the screening of less frequent mutations (such as *BRAF* D594V) alongside the more frequent variants. Like Pyrosequencing, the sequencing primer can be placed adjacent to the mutation site such that the original amplification product does not have to be that big which is advantageous when handling fragmented DNA. The multiplexing capacity of Sequenom has led to the design of the OncoCarta Panel, which targets 238 simple and complex mutations across 19 genes (Pearce *et al.* 2009). Sequenom has been used in numerous studies of *KRAS* mutation status in CRC using FFPE tissue (Fumagalli *et al.* 2010; Arcila *et al.* 2011) and has been shown to be sensitive to low levels of mutant allele (2.5-10%).
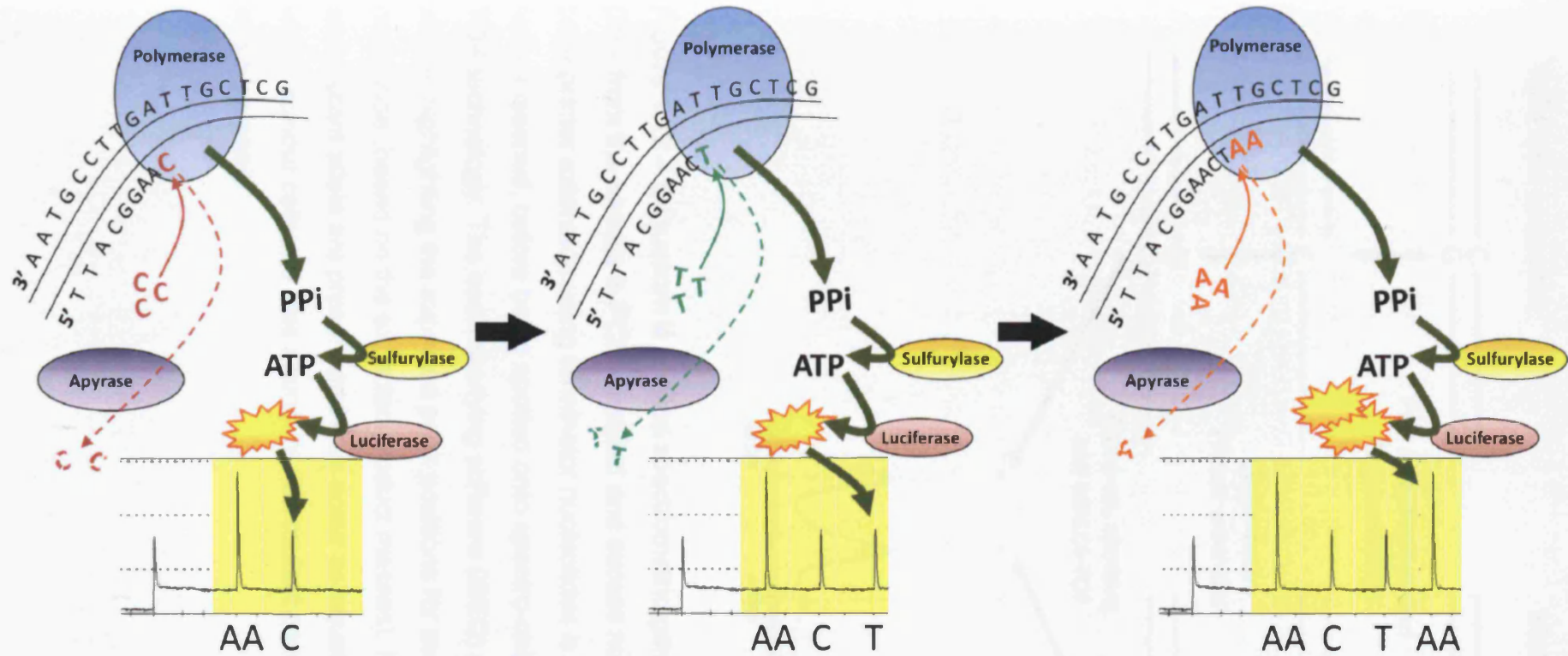
**Figure 1.11** – Pyrosequencing is a 'sequence by synthesis' technology. Nucleotides, added to the reaction in a user-defined order, anneal to a growing synthesis strand. This generates pyrophosphate which is detected as a fluorescent emission, the intensity of which is proportional to number of nucleotides incorporated. Thus if, as depicted in the above diagram a single C anneals, this will generate a fluorescent emission sufficient to generate one C peak on the pyro-diagram. Since T is the next released nucleotide, it will anneal to the A in the template strand and generate a fluorescent emission sufficient to produce a T peak. In the final annealing reaction shown, there are two adjacent Ts in the template strand. Thus, two A nucleotides will anneal in turn generating a fluorescent emission that is twice the intensity of the previous two examples and a peak that is twice as high.
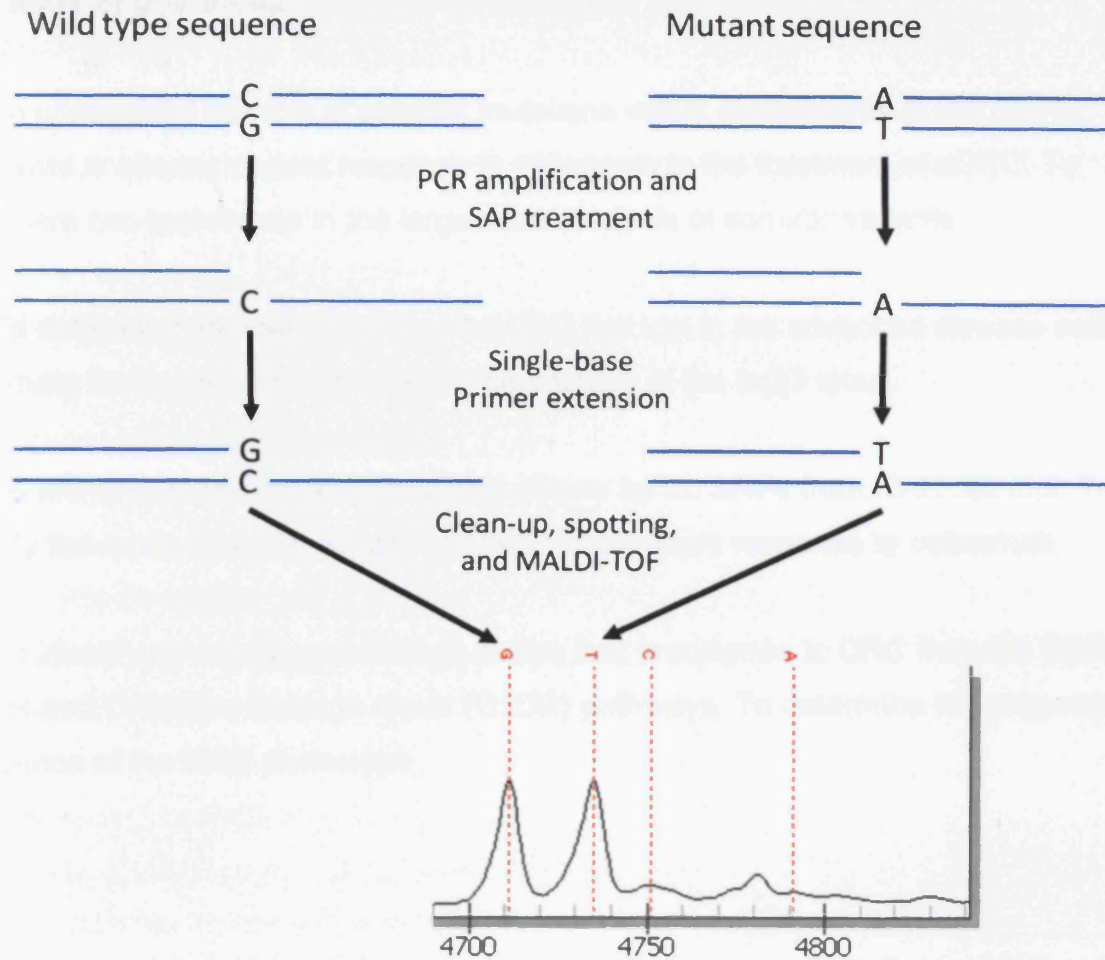
**Figure 1.12** – Sequenom is a mass spectrometric genotyping technology. Genomic DNA from the sample is PCR amplified and excess nucleotides are removed. Single-base primer extension using terminator nucleotides is then performed. Products are further cleaned, before being spotted onto spectro-chips for analysis using MALDI-TOF technology. The accompanying software (iSEQ) provides an output like the one above, highlighting the expected peak positions for every possible mutation at that nucleotide (based on the expected product masses). Here, equal levels of wild-type and mutant allele are present and thus occur as equally sized peaks. In samples where tumour cells are less abundant, the mutant peak will be smaller relative to the wild-type peak.

## 1.7 Aims of this thesis

1, To understand the role of somatic mutations within components of the EGFR pathway in altering patient response to cetuximab in the treatment of aCRC. To compare two techniques in the large scale analysis of somatic variants.

2, To determine the role played by the CRC risk loci in the advanced disease setting. To study the underlying tumorigenic mechanism at the 8q23 locus.

3, To identify prognostic and predictive effects for 20 SNPs from CRC risk loci. To clarify the mechanism by which the 16q22 locus alters response to cetuximab.

4, To identify novel high penetrance alleles that predispose to CRC from the BER, MMR and Oxidative damage repair (OxDR) pathways. To determine the prognostic influence of the MAP phenotype.

# Chapter Two - Materials and Methods

## 2.1 Suppliers

Consumables and equipment used throughout this study were purchased from the following companies:

ABgene Ltd (Surrey, UK)

AGOWA (Berlin, Germany),

Ambion (See Applied Biosystems),

Anachem Ltd (Bedfordshire, UK),

Applied Biosystems (Cheshire, UK),

BD Diagnostics (Franklin Lakes, NJ, USA)

Biorad (Hertfordshire, UK),

Cell Signalling Technologies (Danvers, MA, USA),

Clontech (CA, USA)

Eppendorf (Cambrdgeshire, UK),

Eurogentec (Hampshire, UK),

European Collection of Cell Cultures (Salisbury, UK),

Fisher Scientific (Leicestershire, UK),

Fluka Biochemika (Dorset, UK),

GE Healthcare (Buckinghamshire, UK),

Geneservice Ltd (Nottinghamshire, UK),

IKA (Staufen, Germany),

Illumina (CA, USA),

Imgenex Corp (CA, USA)

Insight Biosight Biotechnology (Santa Cruz, Wembley, UK)

Invitrogen Life Technologies (Strathclyde, UK),

Jencon (See VWR, West Sussex, UK),

Kendro Laboratory Products (Hertfordshire, UK),

Labtech International (East Sussex, UK),

Leica (Wetzlar, Germany),

Lonza Group Ltd (Basel, Switzerland),

Microzone Ltd (Haywards Heath, UK),

Millipore (Hertfordshire, UK),

MJ Research (Massachusetts, USA),

MWG Biotech (Buckinghamshire, UK),

New England Biolabs (Hertfordshire, UK),

Nunc (See Thermo Fisher Scientific),

Olympus Optical (London, UK),

PALM (Bernried, Germany),

Promega (Hampshire, UK),

Qiagen (West Sussex, UK),

Roche Biochemicals (East Sussex, UK),

Sartorius (Surrey, UK),

Santa Cruz Biotechnologies (Santa Cruz, CA, USA),

Sequenom (CA, USA),

Sigma-Aldrich Ltd (Dorset, UK),

Starlab UK Ltd (Milton Keynes, UK),

Starstedt Ltd (Leicestershire, UK),

Stratagene (California, USA),

Thermo Electron Corp (Hampshire, UK),

Thermo Fisher scientific (MA, USA),

VWR International Ltd (West Sussex, UK),

Zeiss (Oberkochen, Germany).


## 2.2. Materials

### 2.2.1 Chemicals

Analytical grade chemicals were purchased from either Sigma-Aldrich Ltd or Fisher Scientific unless otherwise stated.


### 2.2.2 Nucleic acid extraction and purification

QIAamp DNA Micro kits were purchased from Qiagen. Trizol reagent was obtained from Invitrogen Life Technologies. RecoverAll Total Nucleic Acid Isolation Kit was obtained from Ambion Inc.

## 2.2.3 PCR and PCR purification

AmpliTaq Gold DNA polymerase and GeneAmp PCR buffer were obtained from Applied Biosystems. Deoxyribonucleotide triphosphates (dNTPs) were purchased from GE Healthcare. High purity salt free (HPSF) purified-, and HPLC purified biotinylated- oligonucleotide primers were supplied by Eurogentec. Megamix Gold was purchased from Microzone. Exonuclease 1 was purchased from New England Biolabs and Shrimp Alkaline Phosphatase (SAP) as supplied by GE Healthcare respectively. QIAquick PCR purification kits were obtained from Qiagen.

## 2.2.4 Electrophoresis

Multipurpose agarose was purchased from Roche. Ethidium bromide was supplied by Fluka Biochemika whilst the 100bp and 1kb DNA ladders were supplied by New England Biolabs and Invitrogen Life Sciences, respectively.

## 2.2.5 Sequencing

BigDyeTerminator Cycle Sequencing Kit (Version 3.1), POP6 polymer, HiDi Formamide and Genescan 500-ROX size standard were all supplied by Applied Biosystems. Montage SEQ$_{96}$ sequencing reaction clean up kits were obtained from Millipore and capillary electrophoresis running buffers were purchased from Sigma-Aldrich.

## 2.2.6 Antibodies

All antibodies were purchased from Cell Signalling Technology or Santa Cruz Biotechnologies unless otherwise stated.

## 2.2.7 Restriction enzymes

All restriction endonucleases were supplied, along with the appropriate buffer, by New England Biolabs and Pharmacia (Sigma-Aldrich).

## 2.2.8 Cloning

pGEM-T Easy Vector System I, pGL3 and pRL Renilla luciferase vectors and JM109 chemically competent *E.coli* cells were obtained from Promega. Tryptone, yeast extract and agar were supplied by BD Diagnostics Ltd. Ampicillin, X-gal (5-bromo-4-

chloro-3-indoyl-D-galactoside) and IPTG (isopropyl-β-D-thio-galactopyranoside)
were obtained from Sigma-Aldrich Ltd.

## 2.2.9 Clinical Material

All tissue and blood samples were obtained with patient consent and ethical approval
for research in accordance with the guidelines of the COIN Trial.

## 2.2.10 Cell lines and cell line DNA

The HEK293 cell line was kindly provided by Dr. Andrew Tee.

DNA from cell lines used to determine Pyrosequencer and Sequenom assay
sensitivities (CAL-62, AsPC1, RPMI-8226, SW620, DLD1, HCT-116, SW948, Colo-
205, HT117 and HT1197) was kindly provided by various collaborators.

## 2.3 Equipment

### 2.3.1 Plastics and glassware

Sterile pipette tips were supplied by Starlab. RNase- and DNAse-free sterile barrier
tips were obtained from Promega. DistriTips and sterile tips for multi-channel
pipettes were from Anachem. Sterile individually wrapped 5ml, 10ml and 25ml
stripettes were from Corning Costar. Bioquote supplied 0.65ml, 1.5ml and 2.0ml
plastic eppendorfs. Fisher Scientific also supplied 1.5ml eppendorf tubes. The 96
well Thermo-Fast PCR reaction plates were obtained from Thermo Electron
Corporation whilst adhesive PCR sealing sheets, 0.2ml plastic strip tubes and 96
well Thermo-Fast skirted detection plates were purchased from ABgene. Sterile
universals and petri dishes were obtained from Bibby Sterilin and Sarstedt
respectively. Glass flasks and beakers were provided by Jencons or Fisher
Scientific. Optilux 96-well luminometer plates were purchased from VWR
International. Tissue culture flasks were purchased from Nunc. Slides used for
sectioning were purchased from VWR. Fifty millilitre falcon tubes were obtained from
Sarstedt.

## 2.3.2 Laser microdissection

Laser capture microdissection was carried out using the PALM Microlaser system and visualised using the PALM Robo software.

## 2.3.3 Thermocycling

Thermocycling was carried out using an MJ Research DNA Engine Tetrad PTC-225 for PCR.

## 2.3.4 Electrophoresis

DNA electrophoresis was performed using a Horizon 11.14 gel tank from Invitrogen Life Technologies. Visualisation of ethidium bromide stained gels was achieved using a BioRad GelDoc XR transluminator. Power packs were supplied by BioRad.

## 2.3.5 Other equipment

Stained sections of colonic tissue were visualised using a Mirax scanner (Zeiss). Quantitation of DNA and RNA was carried out using the NanoDrop ND-8000 spectrophotometer (Labtech International).

## 2.3.6 Software

Genotyping data from Illumina was viewed using Illumina's GenomStudio program. Statistical analysis and graphing was carried out using Minitab 15 (Minitab Inc.), Microsoft Excel (Microsoft) and SPSS16 (IBM Corp) unless otherwise stated.

When carrying out LCM, slides were visualised using the PALM Robo software (v.1.2.3) (Zeiss).

Assays used for Sequenom analysis were designed using the Massarray Assay Design 3.1 software. Sequenom spectra were analysed using the SpectroREADER software and transferred to the MassARRAY Typer 4 Analyser. Genotyping was performed using the MassARRAY RTTM software. All software was provided through Sequenom.

## 2.4 Methods

### 2.4.1 General reagents

All solutions were made using MilliQ water and autoclaved at 15lb/sq.in at 121°C for 20 minutes where necessary.

1XTAE buffer: 0.04M Tris-acetate, 0.001mM EDTA, pH 8.0

1XTBSbuffer: 0.15M NaCl, 0.005M Tris, pH 7.6

10mM sodium citrate buffer (for 1L – 2.94g sodium citrate trisodium salt dehydrate, 1L $dH_2O$, pH 6.0

### 2.4.2 Sectioning of FFPE tissues

FFPE blocks containing tumour material were placed on ice for at least one hour prior to sectioning. A fresh blade was used to section each block. Ten micrometre sections were cut for LCM use and were placed onto PALM slides. Five micrometer sections were also cut for staining (and visualisation) and macrodissection. Care was taken to ensure that the stained section represented as faithfully as possible the non-stained macrodissected slide (i.e. no excess cutting was performed between sections). Slides carrying sections were incubated at 36°C for normal slides and 60°C for PALM slides overnight.

### 2.4.3 Staining and scanning of sections and identification of tumour material

One of the two 5μm sections was stained using the following protocol; sections were deparaffinised by immersion in xylene twice, 100% ethanol twice, 96% ethanol and 70% ethanol for 2 minutes each. Slides were briefly dipped five times in RNase-free distilled water, stained for 1 minute in Mayer's haematoxylin solution, rinsed for 1 minute in blueing solution and further stained for 10 seconds in Eosin Y. Slides were immersed through an increasing ethanol series for 1 minute each (70%, 95% and 100% ethanol). Sections were then immersed in DPX solution before having a cover slip applied. Slides were then left to dry. Sections used for LCM were stained by the same process but were not immersed in DPX and did not have a cover slip applied. Instead slides were placed back-to-back in a 50ml Falcon tube and immediately frozen (-70°C). This minimalist staining procedure ensured that nucleic acid material could still be obtained from stained tumour material for downstream analysis.

Stained sections were visualised using a Mirax scanner (Zeiss). Slides were scraped to remove excess dry DPX solution. Following this, they were placed into cartridges as groups of 50 for scanning. High-resolution scanning of each slide took ~45 minutes. These scans were subsequently analysed to determine the location of tumour material. Depending on the amount of tumour tissue available, the most appropriate course of tumour extraction (i.e. macrodissection or LCM) was then decided upon. For macrodissection, the paired unstained slide was marked in those areas that carried tumour material.

## 2.4.4 Macrodissection

Extra matching sections were cut from each paraffin embedded block at a thickness of 5μm. These sections were left unstained. Areas containing large and concentrated areas of tumour, identified on the matching stained slide, were drawn around using a marker pen on the underside of the slide. Sterile razor blades were then used to scratch off the tissue lying within the depicted area. Scrapings were emptied into a 1.5ml eppendorf tube for later extraction.

## 2.4.5 Laser capture microdissection (LCM)

Laser micromanipulation provides microscopic high-resolution control of sample composition by enabling the selection or rejection of user defined areas. For LCM, paraffin embedded tissues were sectioned and stained as detailed above. LCM was carried out using a Zeiss Axiovert S100. A pulsed ultra-violet laser is interfaced into the microscope and focused through an objective. The laser cuts the tissue without the heating of adjacent material and results in a clear-cut gap between the desired sample area and the surrounding tissue. After microdissection, the isolated specimens are ejected out of the object plane and catapulted directly into the cap of a 0.5ml eppendorf tube containing 1μl of mineral oil positioned above the slide.

## 2.4.6 DNA extraction from FFPE tissue

DNA was extracted from FFPE tissue using the QIAamp DNA micro kit for both macrodissected and LCM tumour material. Fifteen microlitres of buffer ATL (contents trade secret, CTS) and 10μl of proteinase K were carefully placed into the tube (or lid for LCM). Tubes were pulsed vortexed for 15 seconds. Following this,

56

tubes were left to incubate at 56°C for at least 16 hours. After the tissue was digested, 25μl of buffer ATL was added to the lysate followed by 50μl of buffer AL (CTS) and the tube vortexed for 15 seconds. Fifty microlitres of 100% ethanol was added and the solution was applied to a QIAamp spin column and centrifuged at 13,000 rpm for 1 minute. The column was transferred to a clean collection tube and 500μl of wash buffer AW1 (CTS) was added. The column was re-centrifuged at 13,000 rpm for 1 minute. The eluate was discarded before a second wash was carried out using 500μl of buffer AW2 (CTS). The column was re-centrifuged at 13,000 rpm for 3 minutes followed by an extra 1 minute spin into an empty collection tube to remove residual ethanol. DNA was eluted into ~40μl of DNAse free water by incubating for 1 minute at room temperature and centrifuging at 13,000 rpm for 1 minute into a clean collection tube. Samples were stored at -20°C.

## 2.4.7 RNA extraction from FFPE tissue and conversion to cDNA

Cut sections and LCM cuttings were placed into 1.5ml eppendorfs and immersed in 1ml xylene and mixed for 3 minutes at 50°C on a Thermomixer. Tubes were then centrifuged briefly and the xylene discarded. Remaining pellets were washed twice in 1ml 100% ethanol before being air dried to remove residual ethanol. Digestion buffer and protease (100μl and 4μl respectively) were added to each sample before incubation at 50°C for 15 minutes then at 80°C for 15 minutes. Following this 120μl of isolation additive plus 275μl of 100% ethanol were added to, and mixed with, each sample. Sample mixtures were then passed through a filter cartridge and washed in 700μl of wash 1 (CTS) and 200μl of wash2/3 (CTS) with flow-through being discarded at the appropriate stages. In order to isolate RNA, a mixture of 6μl 10X DNase buffer, 4μl DNase and 50μl nuclease-free water was added to each sample and left to incubate for 30 minutes at room temperature. The washes with wash 1 and wash 2/3 were repeated prior to elution of the RNA with 60μl water at room temperature.

RNA was converted to cDNA using Applied Biosystems High-Capacity cDNA Reverse Transcription Kit. Briefly a mix of 10x RT buffer, 25x dNTP mix (100mM), 10x RTrandom primers, Multiscribe reverse transcriptase, RNase inhibitor and nuclease-free water was prepared on ice. Ten microlitres of each RNA samples was

added to 10μl of the above mix, gently vortexed and centrifuged to bring the tube contents to the bottom.

## 2.4.8 Extraction of DNA from blood samples

DNA was extracted using Qiagen's QIAamp DNA micro kit. One-hundred microlitres of whole blood was transferred into a 1.5ml eppendorf. One hundred microlitres of buffer ATL, 10μl proteinase K and 100μl buffer AL were then added respectively and the tube pulse-vortexed for 15 seconds. Tubes were incubated at 56°C for 10 minutes. To each tube, 50μl of 100% ethanol was added prior to it being vortexed for 15 seconds and incubated at room temperature for 3 minutes. The entire lysate was subsequently transferred to a QIAamp MinElute column and centrifuged at 8,000 rpm for 1 minute with the eluate collected in a collection tube. The column was then transferred to a new collection tube, 500μl buffer AW1 added to the column and then centrifuged at 8,000 rpm for a further 1 minute. This step was repeated with 500μl of buffer AW2. Following this, the tube was centrifuged at 13,000 rpm for 3 minutes to dry the membrane and remove ethanol. The column was placed in a fresh collection tube and 40μl of distilled water added. This was left to stand for 1 minute before a final centrifugation step at 13,000 rpm for 1 minute. DNA samples were stored at -20°C.

## 2.4.9 Quantification of nucleic acids

The concentration of eluted DNA and RNA samples was achieved using ultraviolet (UV) spectrophotometry at wavelengths of 260nm and 280nm to determine the amount of DNA or RNA present and establish endogenous protein content. An absorbance ratio of 1.8 at 260nm:280nm was used as an indicator of high sample purity.

## 2.4.10 Polymerase chain reaction (PCR) and RT-PCR

PCR allows the specific *in vitro* amplification of a defined DNA target sequence in an exponential manner. Double stranded DNA templates are heat denatured and primers bind specifically to complementary target sites on each strand. Thermostable DNA polymerases extend the primers in the 5' to 3' direction by incorporating dNTPs

to create a complementary DNA strand. This cycle is repeated 20 – 40 times enabling newly synthesised DNA molecules to be used as templates at each new round (Mullis *et al.* 1986).

Complimentary primer pairs were designed to have melting temperatures within 2°C of each, to be between 15 and 25 nucleotides in length, lack repetitive motifs and with little predicted dimerization or secondary structure formation.

For standard PCR, 25ng template DNA, 0.25mM dNTPs, 25pmol forward and reverse primers, 2μl 10X GeneAmp PCR buffer (100mM Tris-HCl, pH 8.3, 500mM KCl, 15mM MgCl$_2$, 0.01% w/v gelatin), and 0.5U AmpliTaq Gold DNA polymerase were used in a total volume of 20μl. Cycling conditions were 95°C for 12 minutes, followed by 35 cycles of 95°C for 30 seconds, annealing temperature (52°C - 60°C) for 30 seconds, 72°C for 30 seconds and a final elongation step of 72°C for 10 minutes.

For PCR amplification of lower quality LCM extracted DNA, Megamix Gold (Microzone Ltd.) was used in place of AmpliTaq Gold DNA polymerase. Half of the final reaction volume was made up of the Megamix Gold reagent (CTS), with 25ng template DNA, 25pml forward and reverse primers and water making up the remaining volume. Cycling conditions were the same as described above.

A 20μl RT-PCR reaction contained 1ng cDNA, 0.25mM dNTPs, 25pmmol forward and reverse primers, 2μl 10X GeneAmp PCR buffer (100mM Tris-HCl, pH 8.3, 500mM KCl, 15mM MgCl$_2$, 0.01% w/v gelatine), and 0.5U AmpliTaq Gold DNA polymerase. Cycling conditions were 95°C for 12 minutes, followed by 35 cycles of 95°C for 30 seconds, 58°C for 30 seconds, 72°C for 30 seconds and a final elongation step of 72°C for 10 minutes.

### 2.4.11 Agarose gel electrophoresis

Agarose gels were prepared with 1X TAE buffer to various concentrations between 1.25-2%. Ethidium bromide (0.05μg/ml) was added to the gel to allow for DNA visualisation since it is a DNA intercalating agent that fluoresces under UV light at a

wavelength of 300nm. Three microlitres of loading dye (15%w/v ficol, 10mM Tris pH 8, 1mM EDTA, 0.2% Orange G) was added to each sample before loading and electrophoresis was performed in 1X TAE buffer at 100V. 1kb or 100bp DNA ladders were used to predict fragment sizes. DNA was visualised on a UV transilluminator and photographed using the Bio-Rad XR system.

## 2.4.12 Restriction digest

Restriction digests were performed using enzymes from New England Biolabs/Pharmacia at 1 or 2 units per digest, in the appropriate buffer. Digests were carried out using 10μl of PCR product, run at 37 or 60°C overnight and visualised on an agarose gel.

## 2.4.13 PCR purification

PCR products were purified using an ExoSap method. In a 5μl reaction, 3μl of PCR product was combined with 10U exonuclease I and 0.5U SAP. The sample was incubated at 37°C for 15 minutes to allow for digestion of excess primers and removal of phosphates groups from dNTPS before denaturation of the enzymes at 80°C for 15 minutes.

## 2.4.14 Cycle sequencing and purification

Sanger sequencing uses dideoxynucleotide triphosphates (ddNTPs) which lack the 3' hydroxyl group present in deoxyribose sugars. As a result of this, ddNTPs can be efficiently incorporated into a nascent strand by DNA polymerases but prevent further extension of the growing chain (Sanger et al. 1977). In automated sequencing the reaction can take place within a single tube because each ddNTP is labelled with a different fluorophore. The template DNA is denatured and bound by a single specific primer. DNA polymerase extends this primer by incorporating either an unlabelled dNTP or chain terminating ddNTP at each position. The relative concentrations of dNTPs and ddNTPs are such that the labelled products formed differ in size by one nucleotide. Capillary electrophoresis is used to separate the single stranded DNAs, with smaller fragments migrating fastest through the polymer and passing through the laser beam first. The emitted wavelength of light is detected

and used to determine the ddNTP incorporated at a particular position. The order of the nucleotides provides a sequence read of up to ~500bp.

Sequencing reactions were performed using the BigDye Terminator Cycle Sequencing kit (Version 3.1) according to the manufacturer's instructions. A total reaction volume of 10μl contained 0.6μl – 1.5μl purified PCR product (~5ng), 1μl Terminator ready Reaction Mix (labelled ddNTPs and dNTPs, AmpliTaq DNA polymerase FS, MgCl$_2$ and Tris-HCl buffer, pH 8), 1.6pmol primer and 1.5μl BigDye terminator buffer (CTS). Cycle sequencing conditions were 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 3 minutes and 30 seconds.

The Montage SEQ$_{96}$ Sequencing Reaction Cleanup Kit was used to purify sequencing reactions. Twenty microlitres of injection solution (CTS) was mixed with the reaction and transferred to a filter plate. A vacuum was applied until the wells were empty, then a further 25μl injection solution was added and the vacuum applied again to ensure that all the contaminating salts and unincorporated dyes terminators were filtered out. Purified sequencing products were re-suspended in 25μl injection solution by shaking for 6 minutes. Samples were run on either an ABI 31000 or ABI 3730 Genetic Analyser.

### 2.4.15 Rapid amplification of cDNA ends (RACE) of PCR products

RACE was performed using human colonic Marathon-ready cDNA (Clontech) – pre-prepared libraries of adaptor-ligated cDNA. By using a primer specific to the target sequence of interest, in combination with a primer specific to the adaptor sequence, it is possible to amplify the transcript of interest. Downstream applications such as sequencing allow determination of the 5' and/or 3' ends of the cDNA (named 5'RACE and 3'RACE respectively). Primers used in RACE reactions are required to be 23-28 nucleotides in length, have a GC content of 50-70%, and have a melting temperature of greater than or equal to 65°C. For RACE amplification 36μl water, 5μl 10X cDNA PCR reaction buffer, 1μl dNTP mix (10mM) and 1μl Advantage 2 polymerase mix (50X) are mixed by vortexing for each reaction. This mix is added to a separately prepared mix of 5μl Marathon-ready cDNA, 1μl of the appropriate adaptor primer (10μM) and 1μl of the appropriate gene specific primer (10μM). Cycling conditions were 94°C for 30 seconds, five cycles of 94°C for 5 seconds,

72°C for 4 minutes, five cycles of 94°C for 5 seconds, 70°C for 4 minutes and twenty-five cycles of 94°C for 5 seconds, 68°C for 4 minutes.

RACE products were analysed on an agarose gel by electrophoresis, and sequenced as described above.

## 2.4.16 Real-time PCR

As the name suggests, this technique allows precise real-time monitoring of PCR products as they are generated, through the use of fluorescent probes. Targeted Taqman assays were purchased from Applied Biosystems and used in gene expression analysis, along with Taqman Universal PCR master mix. The human *B2M* gene was used as an endogenous control for all reactions. One hundred nanograms of cDNA was added to each gene expression reaction in a total 20μl volume. Thermal cycling and expression analysis were performed using Applied Biosystems 7500 Fast Real-Time PCR system. Thermal cycling conditions were an initial step of 50°C for 2 minutes and 95°C for 10 minutes, followed by forty cycles of denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 1 minute. Each gene expression reaction was performed in triplicate (for both target and endogenous control assays). PCR specificity of the assay was tested by running the expression product on a 1.5% agarose gel to ensure that only one band was visible. Calibration curves were generated for all samples used and PCR efficiencies calculated to ensure levels between 90-110%.

## 2.4.17 MSI analysis

The microsatellite status of CRC tumour DNA was determined using the BAT25 and BAT26 markers. Amplification of these was carried out using specifically designed primers. Sequences for these were; BAT25F (5' FAM labelled – TCGCCTCCAAGAATGTAAGT, BAT25R – TCTGCATTTTAACTATGGCTC, BAT26F (5' HEX labelled) – TGACTACTTTTGACTTCAGC, BAT26R – AACCATTCAACATTTTTAACC.

The MSI test was standardised using DNA extracted from HCT-166 and Calu1 (MSI and MSS [microsatellite stable] respectively) cell lines. Tumour DNA was serially diluted to test the sensitivity of the assay. Amplification of BAT products was

performed in 12μl reaction volumes containing 4μl of H2O, 6μl of Megamix Gold, 10 pmol of each primer, and 20-30ng of tumour DNA. Thermocycling was performed at 95°C for 10 minutes, followed by thirty-five cycles of 95°C for 1 minute, 60°C for 1 minute and 72°C for 1 minute, followed by a final extension of 72°C for 7 minutes. One microlitre of PCR product was diluted in 180μl $H_2O$ and this was used for subsequent analysis. Microsatellite analysis was carried out using Applied Biosystem's 3100 genescan software, with the microsatellite signal standardised against a ROX marker diluted in Hi Dye Formamide.

## 2.4.18 Immunohistochemistry

The avidin-biotin complex (ABC) method was used for immunohistochemical (IHC) analysis. This method utilises the unique properties of the large glycoprotein avidin and the vitamin biotin which have an extremely high affinity for one another. Biotin can in turn be conjugated to a variety of biological molecules such as antibodies, whilst avidin can be labelled with peroxidise or fluorescein. The technique involves three main steps; application of unlabelled primary antibody, application of biotinylated secondary antibody and application of a complex avidin-biotin peroxidise. The peroxidase is then developed by 3,3'-diaminobenzidine (DAB) or other substrates to produce a coloured end product. The main advantage of this method is the amplification of the original antibody signal due to avidin having four binding sites for biotin, therefore amplifying the signal many fold.

Five micrometer thick paraffin sections containing colorectal tumour tissue were deparaffinised and rehydrated by immersion in xylene twice, 100% ethanol twice, 70% ethanol, 50% ethanol and water for 5 minutes each. For antigen retrieval, sections were boiled in 10mM citrate buffer (pH 6.0) for 10 minutes and gently rinsed under running tap water. Endogenous peroxidase activity was blocked with 0.3% hydrogen peroxide for 10 minutes followed by one 1XTBS wash for 5 minutes. Immunostaining was performed in a humidity chamber using the rabbit/goat VECTASTAIN ELITE ABC horse- and goat-radish peroxidise kits respectively. Sections were encircled with a wax ring and blocked with goat normal serum for 20 minutes. Primary antibodies were applied and incubated overnight at 4°C, followed by three 5 minute washes in 1XTBS. A biotinylated secondary antibody was applied and incubated for 30 minutes followed by three 5 minute 1XTBS washes. ABC was incubated for 30

63

minutes followed by three 5 minute washes in 1XTBS. Sections were developed using DAB, counterstained in Gills haematoxylin for 1 minute, and blued in tap water. Sections were finally dehydrated by immersing in 50% ethanol, 70% ethanol and 100% ethanol twice (all 5 minutes each), cleared in xylene for 10 minutes, mounted with DPX and air dried. Slides were viewed with an Olympus BX51 microscope. All incubations were at room temperature unless otherwise stated.

## 2.4.19 Bacteriological methods

### 2.4.19.1 Bacteriological media and solutions

Unless otherwise stated, solutions were sterilised by autoclaving. LB – Broth; 5g Bacto Tryptone, 2.5g yeast extract and 2.5g NaCl in 500ml $dH_2O$. LB agar: 5g Bacto Tryptone, 2.5g yeast extract, 2.5g NaCl and 8g Bacto Agar in 500ml $dH_2O$. Where appropriate, LB medium had ampicillin added to a final concentration of 100µg/ml, IPTG to 0.5mM and 80µg/ml of X-gal. SOC medium – 2g Bacto Tryptone, 0.5g yeast extract, 1ml 1M NaCl, 0.25ml 1M KCl, 1ml 2M $Mg^{2+}$ stock (filter-sterilised) and 1ml 2M glucose (filter-sterilised).

### 2.4.19.2 Ligation reactions

For pGEM-T easy vectors;

Purified PCR product was ligated with the pGEM-T easy vector (Promega) according to the following conditions; 5µl of 2X rapid ligation buffer (T4 DNA ligase), 1µl of pGEM-T easy vector (50ng), 1µl T4 DNA ligase (3 Weiss units/µl) were used in a reaction with an amount of PCR product defined by the following equation;

$$\text{ng of insert} = \frac{\text{ng of vector x kb size of insert}}{\text{kb size of vector}} \times \text{insert vector molar ratio}$$

Ligations were carried out either at room temperature for 1 hr or 4 °C overnight.

For pGL3 vectors;

The region of interest was PCR amplified using primers containing restriction digest sites for subsequent cloning. The pGL3 vector was digested with the appropriate restriction enzymes in order to generate compatible ends for cloning with them.

During digestion, vector DNA was treated with calf intestinal alkaline phosphatase (CIP) to prevent vector closing. Ligation was carried out using 1µl of digested vector, 3µl of digested PCR product, 5µl 2x ligase buffer and 1µl ligase (3U/µl). Reaction mixes were incubated at 15°C for 24 hours.

## 2.4.19.3 JM109 highly competent Cell Transformation

Transformation of JM109 highly competent *E.coli* was mediated via a heat shock process which led to semi-permeabilisation of the cell membrane enabling the uptake of 'naked' DNA molecules from the surrounding environment into the cell.

Two microlitres of each ligation reaction were placed into 1.5ml eppendorf tubes on ice, with 50µl of just thawed JM109 cells subsequently added. After gentle mixing, samples were placed on ice for 20 minutes. Each sample was heat shocked by placing in a 42°C water bath for exactly 50 seconds and incubated on ice for 2 minutes. The cells were incubated at 37°C with agitation at 225rpm for 1.5 hour after adding 950µl of SOC medium. One-hundred mincrolitres of the growing suspension was spread onto LB agar plates containing ampicillin, X-gal and IPTG. The transformation plates were incubated at 37°C overnight (>16 hours).The pGEM-T easy vector carries the *LacZ* gene which encodes for the enzyme β-galactosidase and so it is possible to identify colonies that contain the insert according to their colour. Cells containing vector without the insert will produce β-galactosidase which leads to the formation of blue colonies because of the utilisation of the enzymes' substrate Xgal. Cells that contain the vector with the successful insert will appear white since the insert disrupts the *LacZ* gene and β-galactosidase is not produced.

## 2.4.19.4 Plasmid minipreps

High quantity plasmid DNA was prepared using QIAprep Spin Miniprep Kit according to the manufacturer's instructions. Around 5-10µg of plasmid DNA is generated from 1.5ml overnight LB cultures. The extraction process involves alkaline lysis of cells accompanied by gentle mixing which releases intact plasmid DNA denatures proteins. Neutralisation and adjustment of the conditions to a high salt medium facilitate binding of DNA to the spin column. High salt conditions cause proteins to denature and chromosomal DNA and cellular debris to precipitate whilst plasmid

DNA stays in solution and binds to the silica-gel membrane of the column. The column/membranes are washed to remove trace nucleases, carbohydrates and salts before the DNA elution.

Clones that successfully carried the correct insert were grown in 3ml of liquid LB containing ampicillin (50µg/ml) at 37°C with agitation (170rpm) for ~15 hours. Cells were centrifuged at 13,000rpm for 1 minute and re-suspended in 250µl of Buffer P1 (100µg/ml RNase A, 50mM Tris/HCl and 10mM EDTA). Two hundred and fifty microlitres of Buffer P2 (200mM NaOH, 1% SDS) was added and the tubes were gently inverted 4-6 times to mix. Three hundred and fifty microlitres of Buffer N3 (3M potassium acetate) was then added, the tubes were gently inverted 4 - 6 times and centrifuged at 13,000rpm for 10 minute. The supernatant was added to a QIAamp column and centrifuged at 13,000rpm for 1 minute. The column was washed with 500µl of Buffer PB (CTS) and then centrifuged at 13,000rpm for 1 minute. The column was washed with 750µl of Buffer PE (containing ethanol and CTS) and centrifuged at 13,000rpm for 1 minute. Residual ethanol from Buffer PE was removed by centrifuging the column twice at 13,000rpm for 1 minute. DNA was eluted into 30µl of $H_2O$ after placing the column in a fresh 1.5ml eppendorf tube and centrifuging at 13,000rpm for 1 minute.

### 2.4.20 Tissue culture

HEK293 adherent cells were grown and maintained in Dulbecco's modified eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% pen-strep. When cells reached ≥80% confluence they were passaged according to the following process. Initially cells were briefly washed with 5ml PBS before being washed with 1ml of Trypsin-EDTA for ~1 minute in order to detach them from the side of the tissue culture flask. Cells then had 2ml of Trypsin added and were incubated at 37°C for 1-3 minutes. The trypsinised cells were transferred into two separate falcon tubes in a final volume of 10ml media (DMEM plus 10% FBS) which were subsequently centrifuged at 1,200rpm for 7-10 minutes.

Following centrifugation, the medium was removed leaving a cell pellet at the bottom of the falcon tube. This was re-suspended in a volume of medium suitable for downstream applications, with care taken to ensure cells were sufficiently mixed with

the medium. This mix was evenly divided between the appropriate numbers of flasks and incubated at 37°C with $CO_2$.


## 2.4.20.1 Transient transfection

Cells were seeded into 96 well plates at $2.5 \times 10^4$ cells per well and allowed to reach 50-80% confluence. Media was then aspirated and replaced with 100μl complete growth medium. In parallel to this, 0.5μg of DNA was diluted with 100μl of Opti-MEM I reduced serum media without serum (Invitrogen). For each well, 0.35μl of Lipofectamine LTX reagent (Invitrogen) was added to the DNA:Opti-MEM mix, mixed and incubated at room temperature for 30 minutes. Following this, 20μl of the DNA:Lipofectamine LTX reagent complex was added to each well. The plate was gently agitated and then left to incubate for a further 24 hours before subsequent assaying was carried out.


## 2.4.20.2 Luciferase assay

HEK293 cells were transfected with 100ng of pGL3 (transformed with the insert of interest) and 0.5ng of pRL (both from Promega) in a 96 well plate. The media was removed and cells washed with 50μl of phosphate buffered saline (PBS). This was immediately removed and replaced with 20μl of 20% diluted lysis buffer. The plate containing the cells was centrifuged at 2,000 rpm for 30 seconds. This was stored at -80°C ready for use.


Luciferase assays were carried out using an Applied Biosystems TR 717 Microplate luminometer. A dual-luciferase reporter assay system was used that allowed the comparison of the activities of firefly and *Renilla* luciferases through sequential measurement. The firefly luciferase reporter was measured first by the addition of 50μl of Luciferase Assay Reagent II (LAR II). Following this the reaction was quenched and the Renilla luciferase reaction initiated by the addition of Stop & Glo reagent. Reactions were repeated at least six times to ensure accurate measurement. The relative activities of the two luciferase reactions give an indication of the expression levels in the pGL3 vector.

# Chapter Three - High-throughput somatic profiling of the Ras-Raf-MAPK and PI3K-PTEN-Akt pathways in advanced colorectal cancer and correlations with response to cetuximab

## 3.1 Introduction

Patients vary in their response to chemotherapy. Inherited factors play a significant role in this response, with germline changes in drug metabolism, transport and target genes all being implicated (Dotor *et al.* 2006; Marcuello *et al.* 2004; Kim *et al.* 2008). In recent years, a role for somatic (tumour based) mutations in altering response has been sought and the introduction of therapies that target the EGFR has shed considerable light into this area. EGFR acts as the gate-way for multiple downstream, intracellular signalling pathways including the Ras-Raf-MAPK and PI3K-PTEN-Akt pathways. Through these pathways, EGFR regulates multiple processes including apoptosis, cellular growth, proliferation, differentiation and migration (Woodburn, 1999). Cetuximab is a monoclonal antibody that binds to the extracellular binding domain of EGFR. In doing so, it prevents the ligands EGF and TGF-α binding to the receptor and triggers receptor internalisation; thus inhibiting downstream signalling. Response to cetuximab has been shown to be limited to patients with CRCs wild-type for *KRAS, BRAF, NRAS* and exon 20 mutations in *PIK3CA* (De Roock *et al.* 2010a). Given the high-frequency of CRCs that are *KRAS* mutant (~40-45%), guidelines are now in place that recommend testing of *KRAS* mutation status prior to treatment with anti-EGFR agents (Allegra *et al.* 2009). However, recent data has suggested that not all somatic mutations within specific target genes affect response to cetuximab. For example, patients treated with cetuximab and with G13D in *KRAS* have longer OS and PFS as compared to patients with other *KRAS*-mutated tumours (De Roock *et al.* 2010b).

Here we tested two mutation detection platforms, Pyrosequencing and Sequenom, for high-throughput somatic profiling of the Ras-Raf-MAPK and PI3K-PTEN-Akt pathways in aCRCs from patients in Arms A (oxaliplatin and 5FU-based chemotherapy) and B (chemotherapy plus cetuximab) of the COIN trial. We determined if there was improved response (OS or PFS) to cetuximab in patients wild-type for *KRAS*. We further analysed *KRAS* in combination with *BRAF* and *NRAS* and determined response to cetuximab in patients carrying a mutation in any

or none of these genes. We determined if *PIK3CA* mutations reduced response to cetuximab. We also determined if mutations of any of the above genes carried a prognostic effect.

## 3.2 Materials and Methods

### 3.2.1 The COIN trial

COIN (COntinuous vs. INtermittent therapy) is a MRC funded fully accrued national trial run through the MRC clinical trials unit (CTU). Tim Maughan was the chief investigator of the study. COIN represents the largest trial of the addition of an EGFR targeted monoclonal antibody (cetuximab) to chemotherapy (oxaliplatin and a fluoropyrimidine) in the first line treatment of aCRC.

### 3.2.2 Patient Samples

All 2,445 patients, from 111 centres across the UK and Republic of Ireland, were randomised in a 1:1:1 ratio to receive either continuous oxaliplatin-based chemotherapy (Arm A), continuous chemotherapy plus cetuximab (Arm B), or intermittent chemotherapy (Arm C), in first line treatment. All patients chose between oral capecitabine (OxCap; two thirds of patients) or infusional 5FU (OxFU; one third) as the partner for oxaliplatin (Figure 3.1). Patients had either previous or current histologically confirmed primary adenocarcinomas of colon or rectum, together with clinical or radiological evidence of advanced and/or metastatic disease, or had histologically/cytologically confirmed metastatic adenocarcinomas, together with clinical and/or radiological evidence of a colorectal primary tumour. All patients gave fully informed consent for their samples to be used for bowel research. Samples were collected as FFPE blocks. Results described herein are for patients from Arms A and B of the trial.
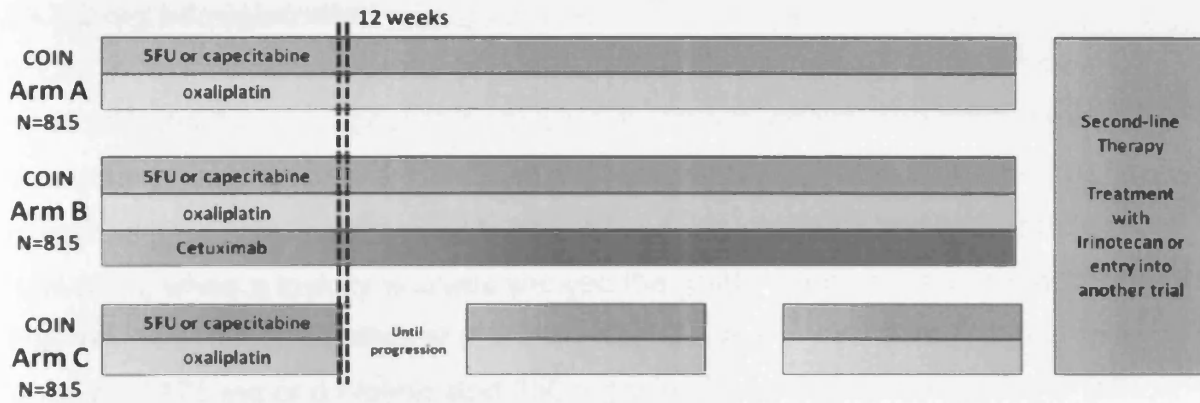
**Figure 3.1** - COIN trial patients were randomised between three equally sized trial arms. Patients in Arm A received a standard chemotherapeutic backbone of OxFU or OxCap, patients in Arm B received the standard chemotherapeutic backbone plus cetuximab, and patients in Arm C received standard chemotherapy but in an intermittent fashion (as part of a separate analysis not considered here). Second-line treatment was irinotecan based therapy or entry into another trial.

### 3.2.3 Drug administration

OxCap was a 3 weekly regimen of intravenous (IV) oxaliplatin 130 mg/m$^2$ over 2 hours followed by oral capecitabine twice daily for two weeks. The initial dose of capecitabine was 1000mg/m$^2$, but was reduced to 850 mg/m$^2$ in a protocol amendment for Arm B patients only after 1775 (73%) patients had been randomised to all arms, when a toxicity analysis showed the grade 3/4 diarrhoea rate was higher than expected (30%) (Adams *et al.* 2009). OxFU was a 2 weekly regimen of IV l-folinic acid 175 mg or d,l-folinic acid 350 mg over 2 hours given concurrently with oxaliplatin 85 mg/m$^2$ over 2 hours, followed by IV bolus 5-FU 400 mg/m$^2$ then 5-FU 2400 mg/m$^2$ infusion over 46 hours administered via an ambulatory pump via a central venous line. In Arm B cetuximab was given in an initial IV dose of 400 mg/m$^2$ over 2 hours and subsequently at 250 mg/m$^2$ over 1 hour once a week.

Treatment was continued until disease progression, development of cumulative toxicity, or patient choice. Patients were allowed to discontinue one or more agents within the regimen for toxicity while continuing on the remaining agent(s).

### 3.2.4 Sample size

The primary analysis of COIN A vs B was planned to take place when 511 OS events (deaths) had been observed among patients wild-type for *KRAS*. In this molecularly selected cohort a higher hazard ratio of 0.76 could be detected at 87% power with a two-sided alpha of 0.05.

### 3.2.5 Processing FFPE CRCs, DNA extraction and MSI analyses

5μm sections were cut from FFPE CRCs using a microtome. One section was stained with Hematoxylin and Eosin and visualised with a Mirax scanner. Samples containing concentrated pockets of tumour material were macrodissected using a second unstained 5μm section. For samples containing limited regions of tumour, LCM was performed using 10μm sections cut onto a PALM membrane slide (Carl Zeiss) and incubated at 56°C for 24 hours. To improve section adherence PALM slides were exposed to UV (254nm) for 30 minutes, incubated in poly-L-Lysine (0.1%w/v) for 5 minutes, and allowed to dry at 60°C for at least 4 hours. Sections were deparaffinised with 100% xylene for 2 minutes (twice), followed by 100%, 95% and 70% ethanol washes for 1 minute. Slides were then dipped 5-6 times in RNase-

free distilled water, stained for 1-2 minutes in Mayer's Hematoxylin solution, rinsed for 1 minute in blueing solution, stained for 10 seconds in Eosin Y, washed in 70%, 95%, and 100% ethanol for 1 minute each and air dried. Slides were viewed and LCM with a Zeiss Axiovert S100 inverted microscope using the PALM Robo software (v.1.2.3). DNA was extracted from macrodissected and LCM tumour material using QIAamp DNA Microkits according to the manufacturer's instruction and eluted in 50µl nuclease-free water. MSI-status was determined using the markers BAT-25 and BAT-26.

### 3.2.6 Identification of somatic mutation 'hot spots' and mutant cell lines

We queried the Catalogue Of Somatic Mutations In Cancer (COSMIC) database (http://www.sanger.ac.uk/genetics/CGP/cosmic) for known common mutations in KRAS, BRAF, NRAS and PIK3CA in CRCs. Cell lines known to carry variants within these genes were identified using the Sanger Cancer Cell Line Project (http://www.sanger.ac.uk/genetics/CGP/CellLines/). We tested the sensitivity of Pyrosequencing and Sequenom to detect low levels of mutant alleles using the cell lines CAL-62 (heterozygous c.34G>C, p.G12R in KRAS), AsPC1 (homozygous c.35G>A, p.G12D in KRAS), RPMI-8226 (homozygous c.35G>C, p.G12A in KRAS), SW620 (homozygous c.35G>T, p.G12V in KRAS), DLD1 (heterozygous c.38G>A, p.G13D in KRAS), HCT-116 (heterozygous c.38G>A, p.G13D in KRAS and heterozygous c.3140A>G, p.H1047R in PIK3CA), SW948 (heterozygous c.182A>T, p.Q61L in KRAS and heterozygous c.1624G>A, p.E542K in PIK3CA), Colo-205 (heterozygous c.1799T>A, p.V600E in BRAF), HT117 (heterozygous c.182A>G, p.Q61R in NRAS) and HT1197 (heterozygous c.1633G>A, p.E545K in PIK3CA). DNA extracted from cell lines was quantified using a nanodrop spectrophotometer and serially diluted with wild-type DNA to generate known levels of mutant compared to wild-type, alleles (50, 25, 12.5, 10, 8, 6, 4, 2 and 0%). All dilutions were prepared and analysed in triplicate.

### 3.2.7 Pyrosequencing

For codons 12 and 13 of KRAS, we initially used the amplification primers 5'-GGCCTGCTGAAAATGACTGA-3' and 5'-AGAATGGTCCTGCACCAGTAATA-3' together with extension primer 5'-CTTGTGGTAGTTGGAGC-3'; however, this assay was subsequently modified by using the extension primers 5'-

TGTGGTAGTTGGAGCTG-3', 5'-TGTGGTAGTTGGAGCT-3' and 5'-
TGGTAGTTGGAGCTGGT-3', as previously described (Ogino *et al.* 2005). For
codon 61 of *KRAS*, we used the amplification primers 5'-
CTTTGGAGCAGGAACAATGTC-3' and 5'-CTCATGTACTGGTCCCTCATTG-3'
together with the extension primer 5'-ATTCTCGACACAGCAGGT-3', and for codon
600 of *BRAF* we used the amplification primers 5'-
TGCTTGCTCTGATAGGAAAATGA-3' and 5'-CAGGGCCAAAAATTTAATCAGTG-3'
together with the extension primer 5'-ATTTTGGTCTAGCTACA-3'. Reverse primers
were biotinylated and purified by HPLC. All other primers were unmodified and
purified by standard SePOP desalting. PCR was performed in 50µl reaction volumes
containing 25µl Megamix Gold, 10-20ng DNA and 10µM of primers. Thermocycling
was performed at 95°C for 10min, followed by 38 cycles of 95°C for 30s, 57°C for
30s, and 72°C for 1min, followed by a final extension of 72°C for 10min. 10µl of each
PCR product was run on a 1.2% agarose gel to confirm amplification, while the
remaining 40µl was used for Pyrosequencing (Qiagen).

For Pyrosequencing, PCR products were immobilised onto Streptavadin sepharose
beads, which were then captured onto probes using Qiagen's vacuum prep tool.
Probes, with the vacuum maintained, were washed in 70% ethanol, denaturation
solution and wash solution. After removing the vacuum, beads were released into a
sequencing primer/annealing buffer mix. Samples were heated to 80°C for 2 min and
allowed to return to room temperature before being placed into the Pyrosequencer
(PyromarkID 96 well format) and run using pre-defined assay conditions. Pyrograms
were analysed by two independent observers.

### 3.2.8 Sequenom

Two hundred base pairs of sequence upstream and downstream of each mutation
were downloaded from Ensembl to design the genotyping assays using the
Sequenom MassARRAY Assay Design 3.1 software. In total, three multiplex assays
were designed. Details of primer sequences are shown in Table 3.1. Multiplex PCR
was performed in 5µl reaction volumes containing 0.5U of Taq polymerase, 5-10ng
of genomic DNA, 100nM of PCR primers and 500µM of dNTP. Thermocycling was
performed at 95°C for 15 min, followed by 45 cycles of 94°C for 20s, 56°C for 30s
and 72°C for 60s, followed by a final extension of 72°C for 3min. Unincorporated

dNTPs were deactivated using 0.3U of SAP at 37°C for 40min, and primer extension was carried out using 7-14µM of each extension primer, 1U of iPLEX termination mix and 1U of iPLEX enzyme. Reactions were cycled at 94°C for 30s, followed by 40 cycles of 94°C for 5s, 52°C for 5s and 80°C for 5s, followed by a final extension at 72°C for 3 min. After the addition of a cation exchange resin to remove residual salt from the reactions, 20µl of water was added and the extension product was spotted onto a matrix pad (3-hydroxypicolinic acid) of a SpectroCHIP (Sequenom). After analysing the SpectroCHIPs using a Bruker MALDI-TOF mass spectrometer, spectra were processed by the SpectroREADER software and transferred to the MassARRAY Typer 4 Analyser (Sequenom). Genotyping was performed using the MassARRAY RTTM software (Sequenom). Automated calls were validated by manual review of the raw mass spectra.

| | Mutation | Forward Primer (5'-3') | Reverse Primer (5'-3') | Extension Primer (5'-3') |
|---|---|---|---|---|
| **A S S A Y 1** | KRAS_G35ACT G12D/A/V | ACGTTGGATGCTGTATCGTCA AGGCACTCT | ACGTTGGATGAGGCCTGCTG AAAATGACTG | AACTTGTGGTAGT TGGAGCTG |
| | KRAS_G38ACT G13D/A/V | ACGTTGGATGCTGTATCGTCA AGGCACTCT | ACGTTGGATGAGGCCTGCTG AAAATGACTG | GCACTCTTGCCTA CG |
| | KRAS_A182CGT Q61P/R/L | ACGTTGGATGTGGAGAAACCT GTCTCTTGG | ACGTTGGATGCATGTACTGGT CCCTCATTG | GGATATTCTCGAC ACAGCAGGTC |
| | KRAS_A183CT Q61H/H | ACGTTGGATGTGGAGAAACCT GTCTCTTGG | ACGTTGGATGCATGTACTGGT CCCTCATTG | ATTGCACTGTACT CCTC |
| | BRAF_T1799A V600E | ACGTTGGATGTGATGGGACC CACTCCATCG | ACGTTGGATGTCTTCATGAAG ACCTCACAG | CCCACTCCATCGA GATTTC |
| | NRAS_G34AT G12S/C | ACGTTGGATGGACTGAGTACA AACTGGTGGT | ACGTTGGATGAGTGGTTCTGG ATTAGCTGGA | GTGCGCTTTTCCC AACACCAC |
| | PIK3CA_G1624A E542K | ACGTTGGATGGCTCAAAGCAA TTTCTACAC | ACGTTGGATGACTTACCTGTG ACTCCATAG | ATTTCTACACAGA TCCTCTCTCT |
| | PIK3CA_G1633A E545K | ACGTTGGATGGCTCAAAGCAA TTTCTACAC | ACGTTGGATGACTTACCTGTG ACTCCATAG | TCCATAGAAAATC TTTCTCCTGCT |
| | PIK3CA_A3140GT H1047R/L | ACGTTGGATGTGAGCAAGAG GCTTTGGAGT | ACGTTGGATGCCAATCCATTT TTGTTGTCC | GAAACAAATGAAT GATGCAC |
| **A S S A Y 2** | KRAS_G34ACT G12S/R/C | ACGTTGGATGCTGTATCGTCA AGGCACTCT | ACGTTGGATGAGGCCTGCTG AAAATGACTG | TGTGGTAGTTGGA GCT |
| | KRAS_G37ACT G13S/R/C | ACGTTGGATGAGGCCTGCTG AAAATGACTGA | ACGTTGGATGGCTGTATCGTC AAGGCACTCT | CACTCTTGCCTAC GC |
| | BRAF_A1781G D594V | ACGTTGGATGTGATGGGACC CACTCCATCG | ACGTTGGATGTCTTCATGAAG ACCTCACAG | CACTGTAGCTAGA CCAAAA |
| | NRAS_C181A Q61K | ACGTTGGATGGTGGTTATAGA TGGTGAAACCT | ACGTTGGATGTATTGGTCTCT CATGGCACTG | ACAGACTGGATA CAGCTGGA |
| | NRAS_A182C Q61P | ACGTTGGATGTATTGGTCTCT CATGGCACTG | ACGTTGGATGGTGGTTATAGA TGGTGAAACCT | TGGCACTGTACTC TTCT |
| | PIK3CA_A1634CG E545A/G | ACGTTGGATGGCTCAAAGCAA TTTCTACAC | ACGTTGGATGACTTACCTGTG ACTCCATAG | AGAGCCTCTCTCT GAAATCACTG |
| | PIK3CA_C1636A Q546K | ACGTTGGATGCCTGTGACTCC ATAGAAAATC | ACGTTGGATGGAACAGCTCAA AGCAATTTC | TCCATAGAAAATC TTTCTCCT |
| **A S S A Y 3** | KRAS_G35ACT G12D/A/V | ACGTTGGATGGCTGTATCGTC AAGGCACTCTT | ACGTTGGATGAGGCCTGCTG AAAATGACTGAA | CTCTTGCCTACGC CA |
| | KRAS_A182CGT Q61P/R/L | ACGTTGGATGCTCATGTACTG GTCCCTCATTG | ACGTTGGATGGATGGAGAAAC CTGTCTCTTGG | ATTGCACTGTACT CCTCT |
| | BRAF_A1781G D594V | ACGTTGGATGATGGGACCCA CTCCATCGAGATT | ACGTTGGATGTTTCTTCATGA AGACCTCACAG | GACTGTAGCTAGA CCAAAA |
| | NRAS_G34ACT G12S/R/C | ACGTTGGATGAGTGGTTCTGG ATTAGCTGGAT | ACGTTGGATGGACTGAGTACA AACTGGTGGTG | CGCTTTTCCCAAC ACCAC |
| | NRAS_A182CGT Q61P/R/L | ACGTTGGATGGTATTGGTCTC TCATGGCACTG | ACGTTGGATGCAAGTGGTTAT AGATGGTGAAAC | ATCCTGGCACTGT ACTCTTCT |
| | PIK3CA_G1624A E542K | ACGTTGGATGGCTCAAAGCAA TTTCTACACAG | ACGTTGGATGGCACTTACCTG TGACTCCATAG | TCCACACAGATCC TCTCT |
| | PIK3CA_G1633A E545K | ACGTTGGATGGCACTTACCTG TGACTCCATAG | ACGTTGGATGGCTCAAAGCAA TTTCTACACAG | TAGAAAATCTTTC TCCTGCT |
| | PIK3CA_A3140GT H1047R/L | ACGTTGGATGCCAATCCATTT TTGTTGTCCAGC | ACGTTGGATGCTGAGCAAGA GGCTTTGGAGTA | TGTCCAGCCACC ATGA |

**Table 3.1** - Primers used in Sequenom analysis of *KRAS, BRAF, NRAS* and *PIK3CA*. Mutation column refers to the target gene and nucleotide change; for example, *KRAS_G35ACT* primers allowed for the analysis of KRAS at the thirty fifth nucleotide of the transcript. A G to A, C or T substitution would be detected. The corresponding amino acid changes are also listed. The nature of the assay designs meant that some nucleotides were assayed more than once.

### 3.2.9 Sanger Sequencing

Sanger sequencing of codons 12 and 13 of *KRAS* was performed using the primers 5'-AAAAGGTACTGGTGGAGTATTTGA-3' and 5'-CATGAAAATGGTCAGAGAAACC-3', codon 61 of *KRAS* was sequenced using 5'-CTTTGGAGCAGGAACAATGTC-3' and 5'-CTCATGTACTGGTCCCTCATTG-3', and codon 600 of *BRAF* was sequenced using 5'-AACTCTTCATAATGCTTGCTCTG-3' and 5'-TGATTTTTGTGAATACTGGGAAC-3'. Single stranded DNA and excess dNTPs were removed using ExoSAP at 37°C for 30min, followed by 80°C for 15min. Cycle-sequencing was performed using Big Dye Mix at 95°C for 10s, followed by twenty-five cycles of 95°C for 10s, 50°C for 5s, 60°C for 200s. Reactions were purified using the Millipore sequencing clean-up plate and analysed on an ABI Prism 3100 genetic analyzer.

### 3.2.10 Statistical Analyses

Analyses were conducted according to a predefined statistical analysis plan which was approved by the COIN trial management group. All randomised patients were included in the analyses, based on the intention-to-treat principle. All P-values are 2-sided and were not adjusted for multiple testing. Time-to-event curves for analysis of OS and PFS were estimated using the Kaplan-Meier method. Hazard ratios, confidence intervals and P-values were estimated using the log-rank method.

Worst toxicity experienced overall was compared between treatment arms using a chi-squared test. For exploratory interaction analyses a Cox proportional-hazards model was fitted separately for each of the covariates to predict PFS in the all wild-type population. All of the above statistical analyses were carried out by David Fisher.

The chi-square test (with Yate's correction) was used for comparison of genotype groups.

### 3.3 Results

### 3.3.1 Determining the sensitivities of the assays

For Pyrosequencing, we initially designed an assay with a single extension primer to detect mutations at codons 12 and 13 of *KRAS*. However, we found that the

76

mutation G12C was not robustly detected in samples with 25% mutant alleles, so the assay was redesigned to a three extension primer system (Ogino *et al.* 2005). This modified assay robustly detected G12A and G12D in samples with 12.5% mutant alleles, G12C and G13D in samples with 6% mutant alleles and G12V in samples with 2% mutant alleles. Sequenom robustly detected G12C in *KRAS* in samples with 6% mutant alleles, V600E in *BRAF* and Q61R in *NRAS* in samples with 10% mutant alleles, Q61L in *KRAS* and E542K, E545K and H1047R in *PIK3CA* in samples with 6% mutant alleles and G12V in *KRAS* in samples with 4% mutant alleles (Table 3.2, Figure 3.2).

|  | gene | mutation | Cell line used | Mutant Allele (%) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  |  |  | 50 | 25 | 12.5 | 10 | 8 | 6 | 4 | 2 | 0 |
| Pyrosequencing | KRAS | G12A | RPMI-8226 | Y | Y | Y | X | X | X | X | X | X |
| | KRAS | G12C | CAL-62 | Y | Y | Y | Y | Y | Y | X | X | X |
| | KRAS | G12D | ASPC1 | Y | Y | Y | X | X | X | X | X | X |
| | KRAS | G12V | SQ620 | Y | Y | Y | Y | Y | Y | Y | Y | X |
| | KRAS | G13D | DLD1 | Y | Y | Y | Y | Y | Y | X | X | X |
| Sequenom | KRAS | G12C | CAL-62 | Y | Y | Y | Y | Y | Y | X | X | X |
| | KRAS | G12V | SQ620 | Y | Y | Y | Y | Y | Y | Y | X | X |
| | KRAS | Q61L | SW948 | Y | Y | Y | Y | Y | Y | X | X | X |
| | BRAF | V600E | COLO-205 | Y | Y | Y | Y | X | X | X | X | X |
| | NRAS | Q61R | HT1197 | Y | Y | Y | Y | X | X | X | X | X |
| | PIK3CA | E542K | SW948 | Y | Y | Y | Y | Y | Y | X | X | X |
| | PIK3CA | E545K | HT1197 | Y | Y | Y | Y | Y | Y | X | X | X |
| | PIK3CA | H1047R | HCT-116 | Y | Y | Y | Y | Y | Y | X | X | X |

**Table 3.2** - Chart depicting the sensitivity of Pyrosequencing and Sequenom to detect somatic mutations in *KRAS, BRAF, NRAS* and *PIK3CA*. Mutant allele that could be detected with confidence is indicated by a green Y while mutant allele that could not be detected is indicated by a red X.

# G12C
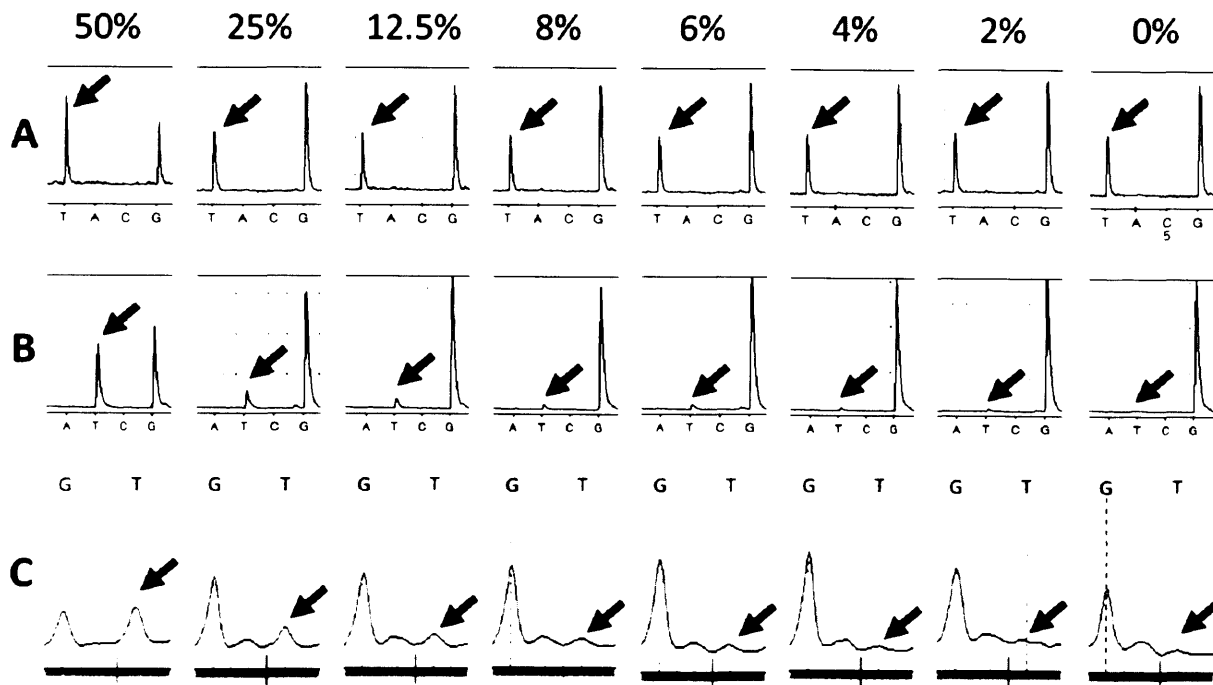


# G12V
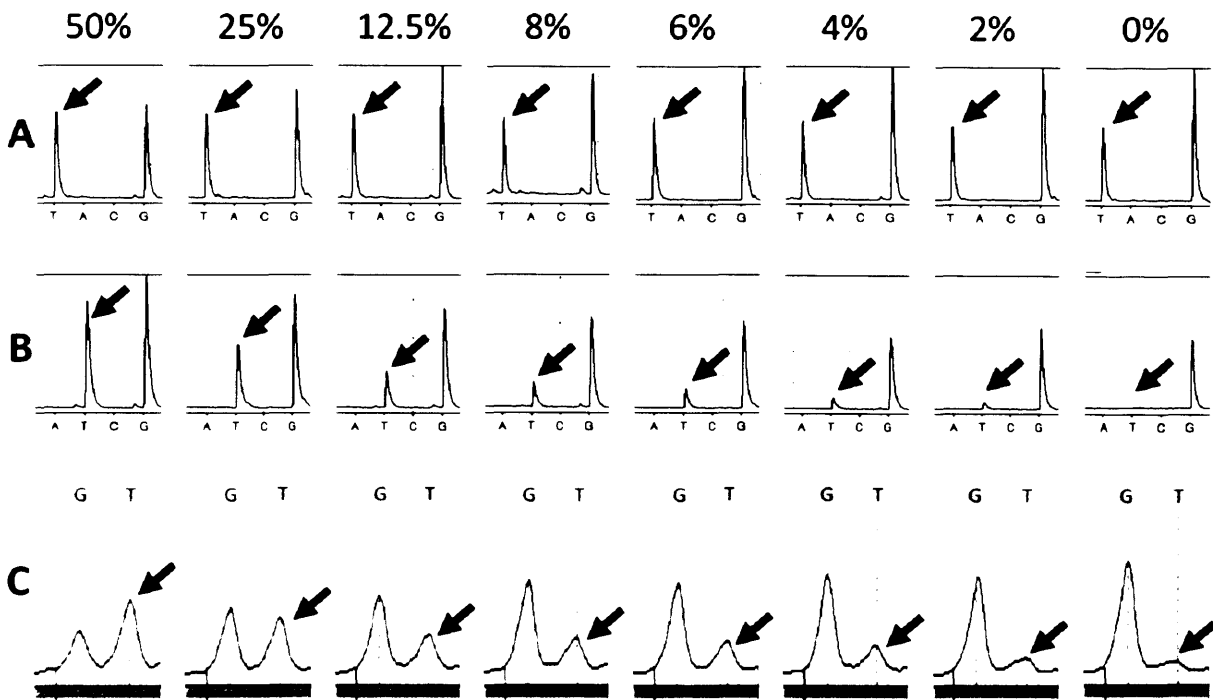


**Figure 3.2** - Diagrams representing detection sensitivities for *KRAS* G12C and G12V using A, Pyrosequencing (single extension primer system), B, Pyrosequencing (multiple extension primer system), and C, Sequenom platforms. Cell lines containing the relevant mutation were diluted with wild-type DNA to concentrations of 50%, 25%, 12.5%, 8%, 6%, 4%, 2% and 0% mutant alleles. Arrows highlight the peak relating to the mutant allele. Traces from A highlight the difficulty in identifying the presence of mutant peaks beyond 12.5% mutant alleles, compared to traces from B and C where *de novo* peaks are generated.

### 3.3.2 Tumour samples

We collected FFPE tumour blocks from 1489/1630 (91.3%) patients from Arms A and B of COIN. One-hundred and seventy-three (11.6%) blocks contained insufficient tumour material for processing. For the remaining samples, 1245 (94.6%) were from the primary CRC, 51 (3.9%) were from liver metastases and 20 (1.5%) were from the lymph metastases.

### 3.3.3 Comparison of Pyrosequencing and Sequenom

We screened for somatic mutations in *KRAS* (codons 12, 13 and 61) and *BRAF* (codon 600) using both Pyrosequencing and Sequenom, and in *BRAF* (codon 594), *NRAS* (codons 12 and 61) and *PIK3CA* (codons 542, 545, 546 and 1047) using just Sequenom. In total, 1,091 samples were successfully analysed for *KRAS* mutations by both technologies and 5803/5860 (99.0%) genotype calls were concordant. For *BRAF* V600E, 884 samples were successfully analysed by both technologies and 869/884 (98.3%) genotype calls were concordant (a breakdown of concordance according to individual nucleotides of *KRAS* and *BRAF* is provided in Table 3.3). Twenty-six out of 57 samples with discordant *KRAS* calls and 8/15 samples with discordant *BRAF* calls were successfully Sanger sequenced to infer genotype. For the remaining calls where Sanger sequencing failed, the mutant genotype was selected (since there was an obvious mutant trace via one technology). Both technologies had high genotype success rates; 1263 of a possible 1308 (96.6%) samples successfully analysed for Sequenom and 1122/1190 (93.6%) for Pyrosequencing.

### 3.3.4 Frequency and distribution of somatic mutations identified

In total, for *KRAS* we successfully genotyped 1294/1316 samples (98.3%), for *BRAF* 1291/1308 samples (98.7%), for *NRAS* 1290/1308 samples (98.6%) and for *PIK3CA* 1256/1308 samples (96.0%). We detected twelve *KRAS* mutations (G12A, G12D, G12V, G12C, G12R, G12S, G13C, G13D, G13V, Q61H, Q61L and Q61R), six *BRAF* mutations (D594G, V600E and four that were uncharacterised), five *NRAS* mutations (G12C, Q61K, Q61L, Q61R and one that was uncharacterised) and five *PIK3CA* mutations (E542K, E545K, Q546K, H1047L, H1047R). Overall, *KRAS* mutations were found in 565/1294 aCRCs (43.7%), *BRAF* mutations in 102/1291 aCRCs (7.9%), *NRAS* mutations in 50/1290 aCRCs (3.9%), and *PIK3CA* mutations in

156/1256 aCRCs (12.4%). Of the *PIK3CA* mutant samples, 107 samples had an exon 9 mutation and 50 samples had an exon 20 mutation (1 sample had a mutation in both exons). In those samples where genotypes were missing for rare mutations (those with cumulative frequencies <1%), but where all other mutations were successfully tested as wild-type, then an overall call of wild-type was made at that locus. The frequencies of individual variants and overlap of mutations across genes are shown in Table 3.4 and Figure 3.3 respectively. MSI was found in 44/969 (4.5%) successfully analysed CRCs. No tumours had mutations in *KRAS* and *BRAF*, although we did find a small number carrying both *KRAS* and *NRAS* mutations (11/565, 1.9%). Mutations in both *KRAS* and *PIK3CA* were frequently observed (90/565 (15.9%)) – analysis of *PIK3CA* mutations within this group revealed that exon 9 mutations significantly co-occurred with *KRAS* mutations (67/565 (11.86%) in *KRAS* mutant versus 40/729 (5.49%) in *KRAS* wild-type; $X^2$=13.583, P=0.0002) but exon 20 mutations did not (24/565 (4.25%) in *KRAS* mutant versus 24/729 (3.29%) in *KRAS* wild-type; $X^2$=0.519, P=0.4711) (Figure 3.4). There was no evidence for any individual or group of mutations of *KRAS* or *PIK3CA* accounting for this significant association. MSI occurred within all mutation sub-groups though there was a significant association between MSI and mutations of *BRAF* (9/102 (8.82%) in *BRAF* mutant versus 35/1189 (2.94%) in *BRAF* wild-type; $X^2$=7.227, P=0.0072) (Figure 3.5).

| Gene | Nucleotide | Mutation | No. samples successfully sequenced by both technologies | No. of calls that agree with consensus | Call success rate (%) | No. incorrect calls (Pyro) | No. incorrect calls (Sequenom) |
|---|---|---|---|---|---|---|---|
| KRAS | 34 | G12C | 1106 | 1099 | 99.37 | 0 | 1 |
| | | G12R | | | | 1 | 0 |
| | | G12S | | | | 1 | 4 |
| | 35 | G12A | 1109 | 1079 | 97.29 | 2 | 0 |
| | | G12D | | | | 9 | 3 |
| | | G12V | | | | 9 | 7 |
| | 37 | G13C | 1102 | 1098 | 99.64 | 4 | 0 |
| | 38 | G13D | 1100 | 1088 | 98.91 | 3 | 8 |
| | | G13V | | | | 0 | 1 |
| | 182 | Q61L | 812 | 811 | 99.88 | 0 | 0 |
| | | Q61R | | | | 1 | 0 |
| | 183 | Q61H | 631 | 628 | 99.52 | 1 | 2 |
| BRAF | 1799 | V600E | 884 | 869 | 98.30 | 14 | 1 |

**Table 3.3** – Table highlighting the sensitivity of Pyrosequencing and Sequenom towards individual mutations of *KRAS* and *BRAF*. Pyro - Pyrosequencing

| Gene | Mutation | Arm A<br>no. of samples with<br>mutations/no.<br>successfully analysed | Arm B<br>no. of samples with<br>mutations/no.<br>successfully analysed | Total<br>no. of samples with<br>mutations/no.<br>successfully analysed |
|---|---|---|---|---|
| *KRAS* | G12A | 23/635 (3.62%) | 11/659 (1.67%) | 34/1294 (2.63%) |
| *KRAS* | G12D | 74/635 (11.65%) | 94/659 (14.26%) | 168/1294 (12.98%) |
| *KRAS* | G12V | 59/635 (9.29%) | 82/659 (12.44%) | 141/1294 (10.90%) |
| *KRAS* | G12C | 23/635 (3.62%) | 14/659 (2.12%) | 37/1294 (2.86%) |
| *KRAS* | G12R | 6/635 (0.94%) | 5/659 (0.76%) | 11/1294 (0.85%) |
| *KRAS* | G12S | 14/635 (2.20%) | 20/659 (3.03%) | 34/1294 (2.63%) |
| *KRAS* | G13C | 3/635 (0.47%) | 2/657 (0.91%) | 5/1292 (0.39%) |
| *KRAS* | G13D | 56/635 (8.82%) | 54/659 (8.19%) | 110/1294 (8.50%) |
| *KRAS* | G13V | 0/635 (0%) | 1/659 (0.15%) | 1/1294 (0.08%) |
| *KRAS* | Q61H | 5/518 (0.97%) | 8/541 (1.48%) | 13/1059 (1.23%) |
| *KRAS* | Q61L | 2/633 (0.32%) | 3/656 (0.46%) | 5/1289 (0.39%) |
| *KRAS* | Q61R | 3/633 (0.47%) | 3/656 (0.46%) | 6/1289 (0.47%) |
| *Total*<br>*KRAS* | | **268/635 (42.20%)** | **297/659 (45.07%)** | **565/1294 (43.66%)** |
| *BRAF* | V600E | 50/631 (7.92%) | 40/660 (6.06%) | 90/1291 (6.97%) |
| *BRAF* | D594G | 7/622 (1.13%) | 5/655 (0.76%) | 12/1277 (0.94%) |
| *Total*<br>*BRAF* | | **57/630 (9.05%)** | **45/661 (6.81%)** | **102/1291 (7.90)** |
| *NRAS* | G12C | 0/621 (0%) | 11/653 (1.68%) | 11/1274 (0.86%) |
| *NRAS* | Q61K | 10/612 (1.63%) | 12/634 (1.89%) | 22/1246 (1.77%) |
| *NRAS* | Q61L | 2/633 (0.32%) | 5/652 (0.77%) | 7/1285 (0.54%) |
| *NRAS* | Q61R | 6/633 (0.95%) | 3/652 (0.46%) | 9/1285 (0.70%) |
| *Total*<br>*NRAS* | | **18/631 (2.85%)** | **32/659 (4.86%)** | **50/1290 (3.88%)** |
| *PIK3CA* | E542K | 20/612 (3.27%) | 22/634 (3.47%) | 42/1246 (3.37%) |
| *PIK3CA* | E545K | 27/613 (4.40%) | 29/636 (4.56%) | 56/1249 (4.48%) |
| *PIK3CA* | Q546K | 3/610 (0.49%) | 10/629 (1.59%) | 13/1239 (1.05%) |
| *PIK3CA* | H1047L | 10/615 (1.63%) | 13/632 (2.06%) | 23/1247 (1.84%) |
| *PIK3CA* | H1047R | 13/615 (2.11%) | 14/632 (2.22%) | 27/1247 (2.17%) |
| *Total*<br>*PIK3CA* | | **71/617 (11.51%)** | **85/639 (13.30%)** | **156/1256 (12.42%)** |

**Table 3.4** – Summary of mutation frequencies across all four genes analysed. Frequencies are displayed for Arms A and B of COIN, and combined. Total numbers per locus to do not exactly match individual numbers since: (i) for *KRAS*, four samples contained two independent mutations and four other samples contained uncharacterised mutations, (ii) for *NRAS*, one sample contained an uncharacterised mutation and, (iii) for *PIK3CA*, five samples contained two independent PIK3CA mutations.
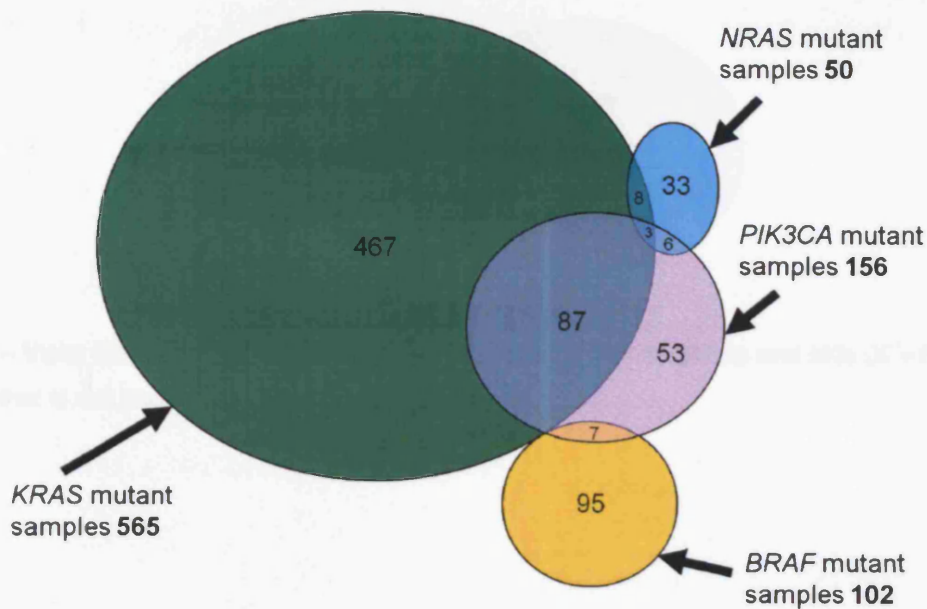
**Figure 3.3** - Venn diagram depicting the overlap between tumours carrying different mutations in our cohort. *KRAS* mutant samples are depicted in dark green, *BRAF* mutant samples in yellow, *NRAS* mutant samples in blue, and *PIK3CA* mutant samples in purple.
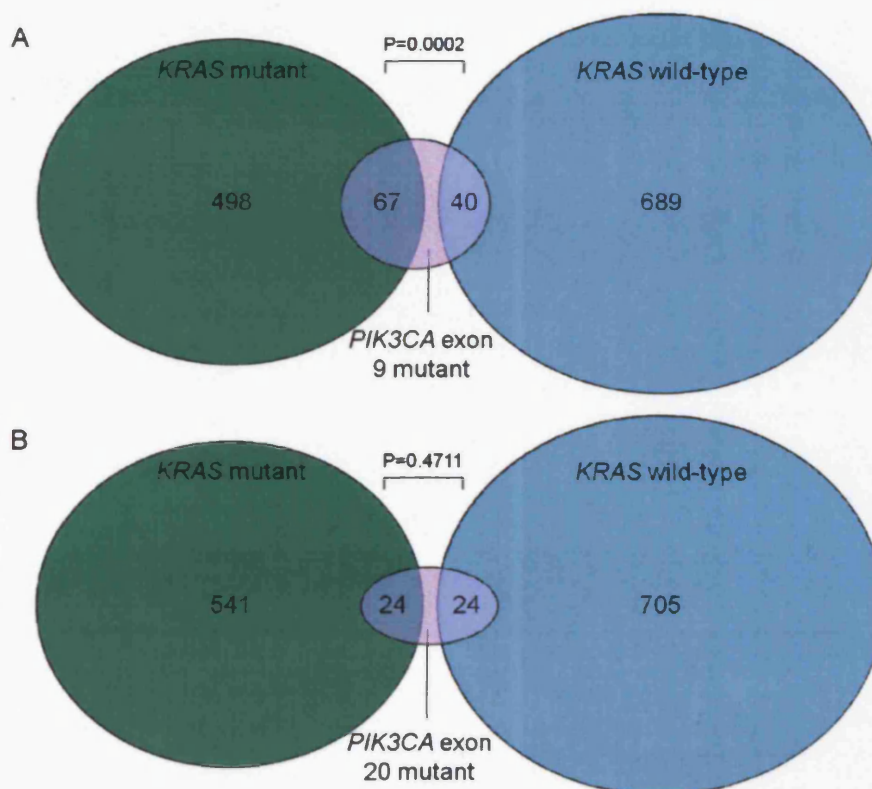


**Figure 3.4** – Venn diagram depicting the overlap between *KRAS* and *PIK3CA* mutations according to exon. (A) There is a significant overlap between *PIK3CA* exon 9 mutations and *KRAS* mutations ($X^2$=13.583, P=0.0002) that is not present in *KRAS* wild-type tumours. (B) There is no significant overlap between *PIK3CA* exon 20 mutations and *KRAS* mutant or wild-type states. One patient had a mutation of both exons.
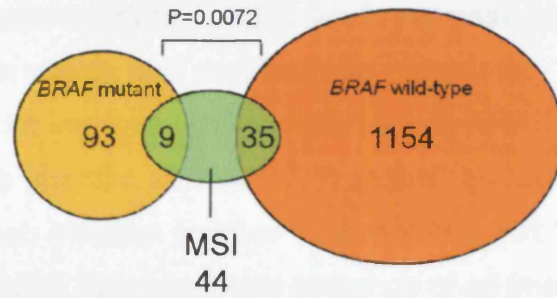
Figure 3.5 – Venn diagram depicting the overlap between *BRAF* mutations and MSI ($X^2$=7.227, P=0.0072) that is not present in *BRAF* wild-type tumours.

### 3.3.5 Analysis of somatic mutation status and response to cetuximab

### 3.3.5.1 *KRAS* mutation status and response to cetuximab

We found no evidence for improved OS in patients with *KRAS* wild-type tumours administered cetuximab (median survival 17.0 months) as compared to those that did not receive cetuximab (median survival 17.9 months) (HR 1.04, 95% CI 0.87-1.23, P=0.67) (Figure 3.6A). Nor was there evidence of an improvement or decline in OS in patients with *KRAS* mutant tumours administered cetuximab (HR 0.98, 95% CI 0.81-1.17, P=0.80) (Figure 3.6B). Similarly for PFS, patients with *KRAS* wild-type (HR 0.96, 95% CI 0.82-1.12, P=0.60) (Figure 3.6C) and mutant tumours (HR 1.06, 95% CI 0.90-1.26, P=0.47) (Figure 3.6D) failed to show a benefit from cetuximab use.

### 3.3.5.2 *KRAS*, *BRAF* and *NRAS* mutation status and response to cetuximab

Analysis of patients wild-type for all of *KRAS*, *BRAF* and *NRAS* also revealed a lack of benefit from cetuximab in terms of OS (median OS 20.1 versus 19.9 months, HR 1.02, 95% CI 0.83-1.24, P=0.86) (Figure 3.7A) and PFS (HR 0.92, 95% CI 0.78-1.10, P=0.36) (Figure 3.7C). There was no improvement or decline in response to cetuximab in patients with a mutation of any of these gene (HR 1.00, 95% CI 0.85-1.18, P=0.96) (Figure 3.7B).

Exploratory analysis of the COIN trial data identified a subset of patients most likely to gain benefit from cetuximab treatment. These were patients that were 'all wild-type' for *KRAS*, *BRAF* and *NRAS*, received OxFU (HR 0.72, P=0.04) instead of OxCap (HR 1.02, P=0.88, P for interaction=0.10) and had 0/1 metastatic sites (HR 0.73, P=0.03) instead of 2 or more (HR 1.07, P=0.56, P for interaction=0.04). This subgroup of patients, deemed the most responsive cohort, showed significant benefit for PFS (n=96, HR 0.55, 95% CI 0.35-0.87, P=0.01) and a trend towards improved OS (HR 0.63, 95% CI 0.38-1.05, P=0.076) (Figure 3.8).
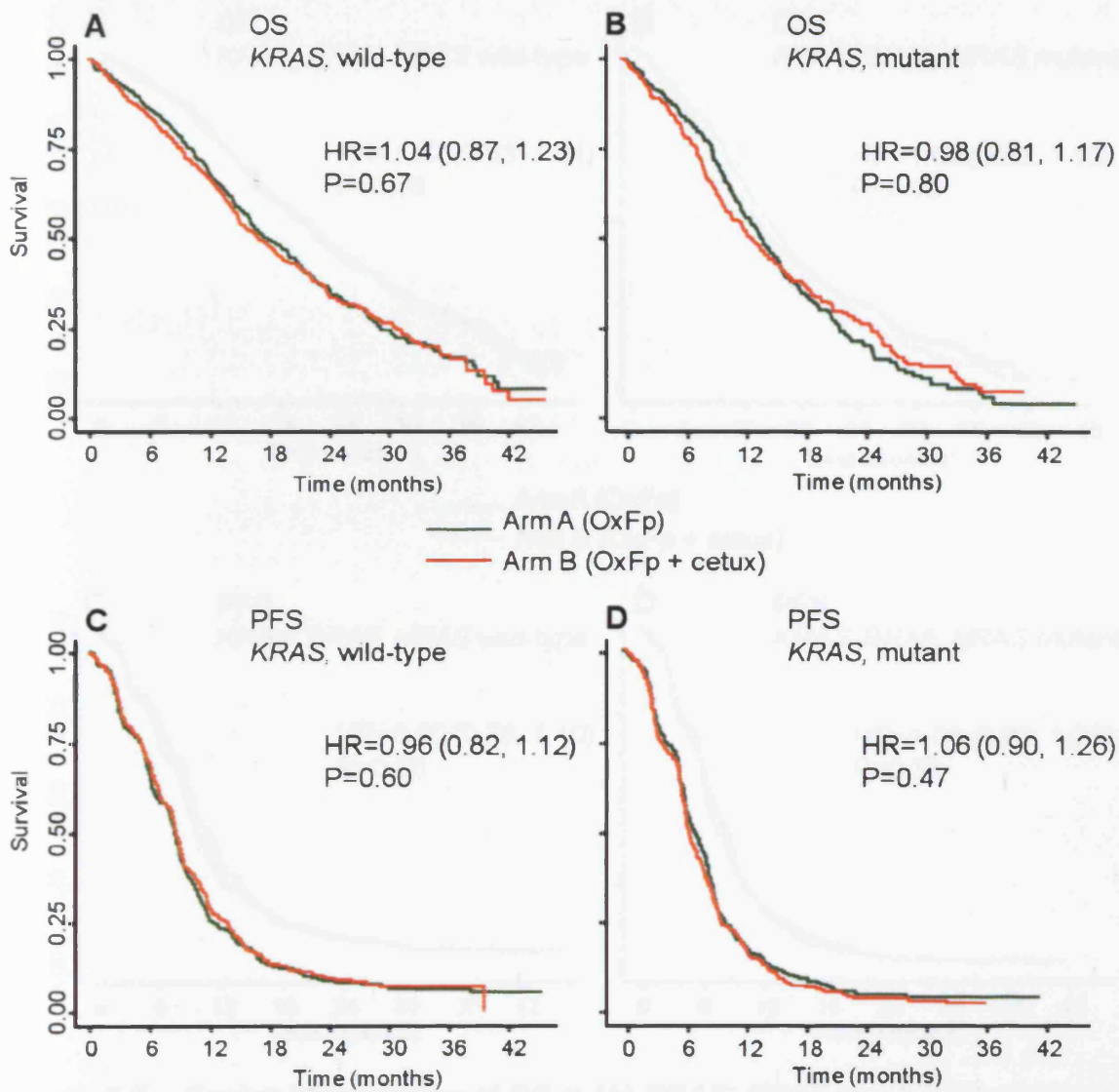
**Figure 3.6** – Kaplan-Meier curves of OS in (A) *KRAS* wild-type, (B) *KRAS* mutant patients, and PFS in (C) *KRAS* wild-type, (D) *KRAS* mutant patients. OxFp indicates that patients received oxaliplatin plus a fluoropyrimidine (5FU or capecitabine).
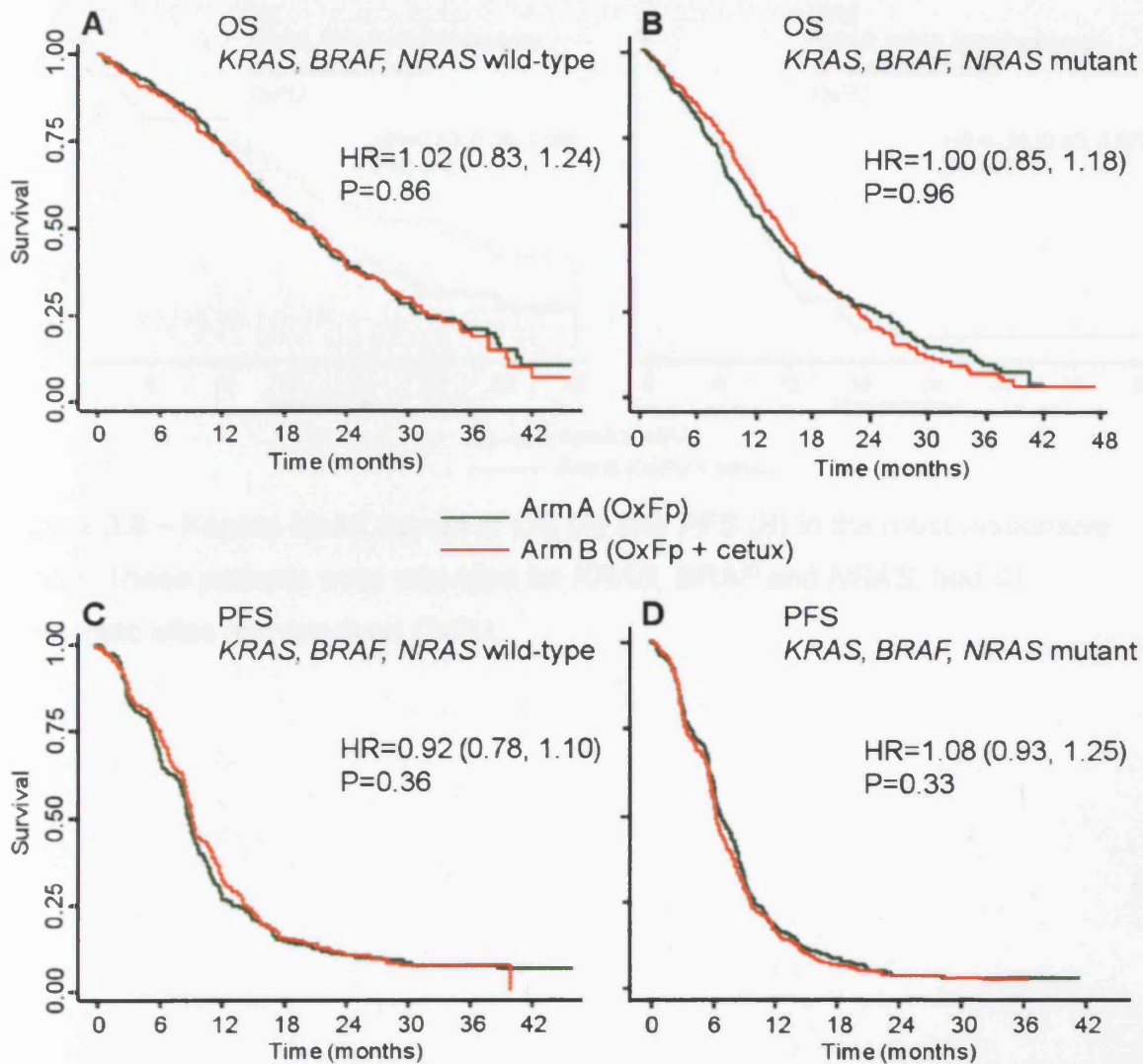
**Figure 3.7** – Kaplan-Meier curves of OS in (A) *KRAS, BRAF* and *NRAS* wild-type, (B) *KRAS, BRAF* and *NRAS* mutant patients and PFS in (C) *KRAS, BRAF* and *NRAS* wild-type samples, (D) *KRAS, BRAF* and *NRAS* mutant patients. OxFp indicates that patients received oxaliplatin plus a fluoropyrimidine (5FU or capecitabine).
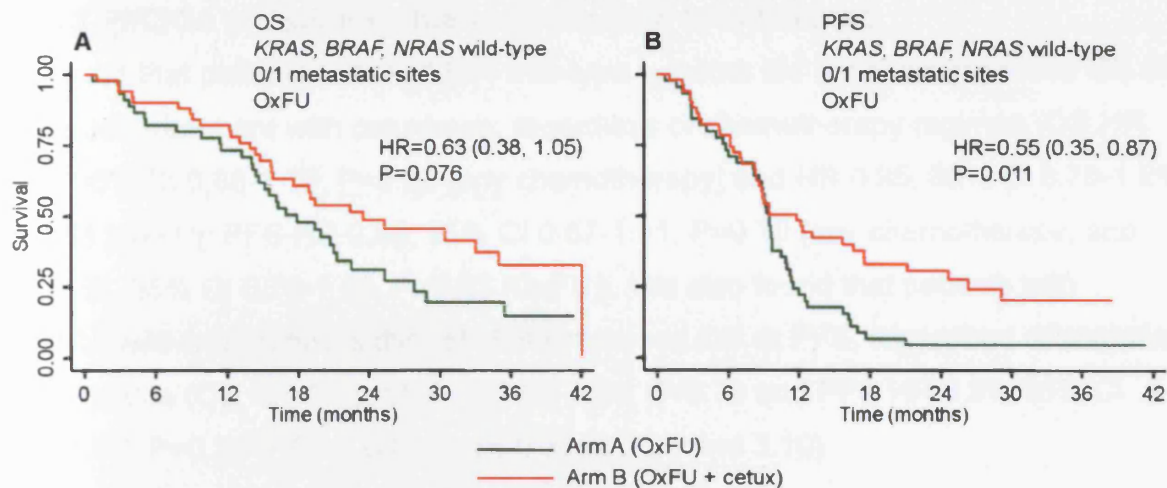
**Figure 3.8** – Kaplan-Meier curves of OS (A) and PFS (B) in the most responsive cohort. These patients were wild-type for *KRAS*, *BRAF* and *NRAS*, had ≤1 metastatic sites and received OxFU.

### 3.3.5.3 *PIK3CA* mutation status and response to cetuximab

We found that patients with *PIK3CA* wild-type tumours did not show improved OS or PFS upon treatment with cetuximab, regardless of chemotherapy regimen (OS HR 1.01, 95% CI 0.88-1.16, P=0.89 [any chemotherapy] and HR 0.95, 95% CI 0.75-1.21, P=0.68 [OxFU]; PFS HR 0.98, 95% CI 0.87-1.11, P=0.79 [any chemotherapy] and HR 0.88, 95% CI 0.70-1.09, P=0.23 [OxFU]). We also found that patients with *PIK3CA* wild-type tumours did not show improved OS or PFS, regardless of somatic *KRAS* status (OS HR 1.03, 95% CI 0.86-1.24, P=0.75 and PFS HR 0.91, 95% CI 0.77-1.07, P=0.25 [*KRAS* wild-type]) (Figures 3.9 and 3.10).

### 3.3.5.4 MSI status in COIN patients

There was a slight suggestion of an adverse effect on survival from cetuximab treatment among patients with MSI, but this did not reach significance due to the very small numbers (n=44; OS HR 1.26, 95% CI 0.64-2.48, P=0.50; PFS HR 1.21, 95% CI 0.65-2.28, P=0.55).

### 3.3.5.5 Individual somatic status and response to cetuximab

We also determined whether any individual mutations were consistently associated with an improvement or a decline in response to cetuximab within the full cohort and in those patients treated with OxFU (Figure 3.11). Several trends were observed, although all bar one were statistically insignificant. As expected, most somatic mutations were consistently associated with a poor response to cetuximab, for example patients with G12D in *KRAS* showed a decline in PFS in response to cetuximab (HR 1.29, 95% CI 0.94-1.77, n=168 [any chemotherapy] and HR 1.42, 95% CI 0.78-2.60, n=48 [OxFU]), as did patients with E542K in *PIK3CA* (HR 1.37, 95% CI 0.73-2.59, n=42 [any chemotherapy] and HR 1.37, 95% CI 0.42-4.48, n=14 [OxFU]). However, patients with G12V in *KRAS* showed a modestly improved PFS in response to cetuximab (HR 0.92 95% CI 0.65-1.31, n=141 [any chemotherapy] and HR 0.78, 95% CI 0.43-1.41, n=55 [OxFU]) as did patients with E545K in *PIK3CA* (HR 0.82, 95% CI 0.47-1.42, n=56 [any chemotherapy], HR 0.75, 95% CI 0.29-1.92, n=20 [OxFU]).
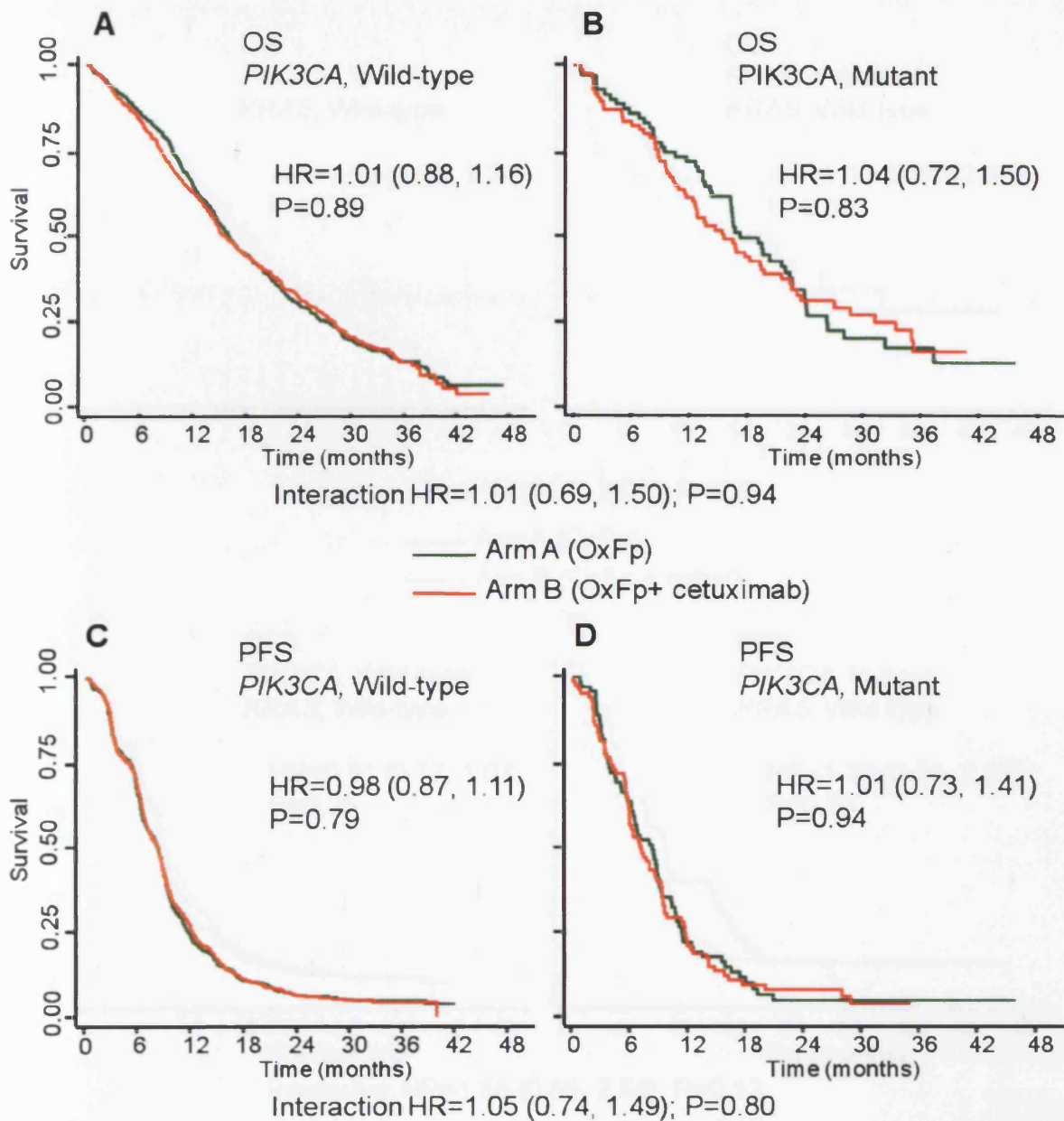
**Figure 3.9** - Kaplan-Meier curves of OS in (A) *PIK3CA* wild-type, (B) *PIK3CA* mutant patients, and PFS in (C) *PIK3CA* wild-type, (D) *PIK3CA* mutant patients. OxFp indicates that patients received oxaliplatin plus a fluoropyrimidine (5FU or capecitabine).
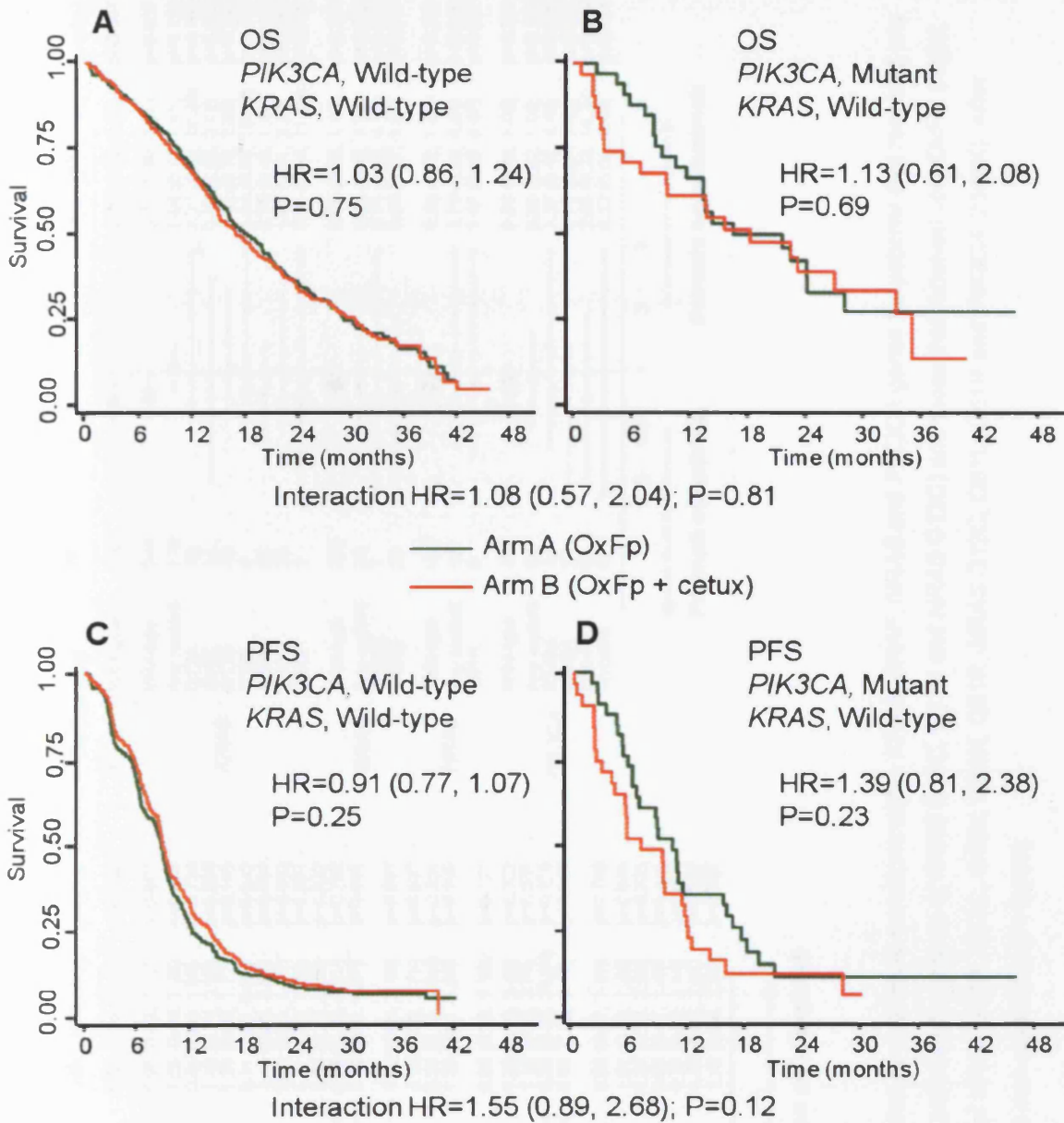
**Figure 3.10** - Kaplan-Meier curves of OS in (A) *PIK3CA* wild-type, *KRAS* wild-type, (B) *PIK3CA* mutant, *KRAS* wild-type patients and PFS in (C) *PIK3CA* wild-type, *KRAS* wild-type, (D) *PIK3CA* mutant, *KRAS* wild-type patients. OxFp indicates that patients received oxaliplatin plus a fluoropyrimidine (5FU or capecitabine).

## PFS

| Subgroup | | N | HR (95% CI) | p-value |
|---|---|---|---|---|
| All | Wild-type | 581 | 0.92 (0.78, 1.10) | p=0.359 |
| KRAS | Wild-type | 729 | 0.96 (0.82, 1.12) | p=0.602 |
| | Any mutant | 565 | 1.06 (0.90, 1.26) | p=0.465 |
| | G12A | 34 | 0.89 (0.42, 1.89) | p=0.754 |
| | G12C | 37 | 1.48 (0.74, 2.98) | p=0.268 |
| | G12D | 168 | 1.29 (0.94, 1.77) | p=0.115 |
| | G12R | 11 | 0.11 (0.01, 0.94) | p=0.043 |
| | G12S | 34 | 1.40 (0.68, 2.92) | p=0.364 |
| | G12V | 141 | 0.92 (0.65, 1.31) | p=0.652 |
| | G13D | 110 | 1.11 (0.76, 1.63) | p=0.587 |
| | Q61H | 13 | 0.66 (0.19, 2.33) | p=0.523 |
| | Q61L | 5 | 0.76 (0.10, 5.51) | p=0.786 |
| | Q61R | 6 | 0.70 (0.11, 4.27) | p=0.695 |
| BRAF | Wild-type | 1189 | 1.00 (0.89, 1.12) | p=0.987 |
| | Any mutant | 102 | 1.14 (0.76, 1.71) | p=0.537 |
| | D594G | 12 | 0.51 (0.13, 2.01) | p=0.335 |
| | V600E | 90 | 1.25 (0.81, 1.94) | p=0.309 |
| NRAS | Wild-type | 1240 | 0.98 (0.88, 1.10) | p=0.777 |
| | Any mutant | 50 | 1.35 (0.75, 2.45) | p=0.322 |
| | Q61K | 22 | 1.12 (0.46, 2.70) | p=0.808 |
| | Q61L | 7 | 3.32 (0.38, 28.94) | p=0.278 |
| | Q61R | 9 | 0.96 (0.21, 4.40) | p=0.957 |
| PIK3CA | Wild-type | 1100 | 0.98 (0.87, 1.11) | p=0.788 |
| | Any mutant | 156 | 1.01 (0.73, 1.40) | p=0.944 |
| | E542K | 42 | 1.37 (0.73, 2.59) | p=0.325 |
| | E545K | 56 | 0.82 (0.47, 1.42) | p=0.478 |
| | Q546K | 23 | 0.25 (0.05, 1.14) | p=0.074 |
| | H1047L | 27 | 1.12 (0.48, 2.61) | p=0.787 |
| | H1047R | 13 | 1.16 (0.50, 2.70) | p=0.727 |

0.25  0.5  1  2  4

◄‑‑‑‑‑‑‑‑ Favours cetuximab      Favours no cetuximab ‑‑‑‑‑‑‑‑►

## PFS; OxFU pts only

| Subgroup | | N | HR (95% CI) | p-value |
|---|---|---|---|---|
| All | Wild-type | 190 | 0.72 (0.53, 0.98) | p=0.037 |
| KRAS | Wild-type | 244 | 0.77 (0.59, 1.01) | p=0.057 |
| | Any mutant | 179 | 1.04 (0.77, 1.41) | p=0.775 |
| | G12A | 10 | 3.27 (0.65, 16.43) | p=0.150 |
| | G12C | 14 | 0.95 (0.29, 3.12) | p=0.930 |
| | G12D | 48 | 1.42 (0.78, 2.60) | p=0.249 |
| | G12S | 9 | 3.54 (0.68, 18.49) | p=0.133 |
| | G12V | 55 | 0.78 (0.43, 1.41) | p=0.414 |
| | G13D | 30 | 1.07 (0.51, 2.24) | p=0.860 |
| | Q61L | 3 | 0.71 (0.04, 11.79) | p=0.809 |
| BRAF | Wild-type | 383 | 0.89 (0.72, 1.10) | p=0.278 |
| | Any mutant | 36 | 0.85 (0.42, 1.73) | p=0.652 |
| | D594G | 7 | 0.74 (0.13, 4.13) | p=0.728 |
| | V600E | 29 | 0.83 (0.37, 1.87) | p=0.647 |
| NRAS | Wild-type | 403 | 0.89 (0.72, 1.09) | p=0.249 |
| | Any mutant | 16 | 1.70 (0.59, 4.94) | p=0.329 |
| | Q61K | 9 | 1.47 (0.36, 5.97) | p=0.587 |
| PIK3CA | Wild-type | 359 | 0.88 (0.70, 1.09) | p=0.231 |
| | Any mutant | 54 | 0.90 (0.52, 1.58) | p=0.725 |
| | E542K | 14 | 1.37 (0.42, 4.48) | p=0.605 |
| | E545K | 20 | 0.75 (0.29, 1.92) | p=0.551 |
| | H1047L | 8 | 2.05 (0.23, 18.24) | p=0.518 |
| | H1047R | 11 | 0.32 (0.08, 1.40) | p=0.131 |

0.25  0.5  1  2  4

◄‑‑‑‑‑‑‑‑ Favours cetuximab      Favours no cetuximab ‑‑‑‑‑‑‑‑►

**Figure 3.11** - Forest plots depicting PFS hazard ratios for individual mutations within *KRAS*, *BRAF*, *NRAS* and *PIK3CA* genes for patients that received any chemotherapy (Left; There was insufficient data to estimate the effects of *KRAS* G13C, G13V and *NRAS* G12C) and those that received only OxFU (Right; There was insufficient data to estimate the effects of *KRAS* G12R, G13C, G13V, Q61H, Q61R, *NRAS* G12C, Q61L, Q61R and *PIK3CA* Q546K). After correction for multiple testing, all observations were non-significant (P>0.999).

It has been suggested that patients with *KRAS* p.G13D-mutated tumours treated with cetuximab have longer OS and PFS as compared to patients with other *KRAS*-mutated tumours. However, we found that patients with G13D treated with cetuximab showed no difference in OS (HR 1.04, 95% CI 0.68-1.59, P=0.86 [any chemotherapy]) or PFS (HR 1.11, 95% CI 0.76-1.63, P=0.59 [any chemotherapy]; HR 1.07, 95% CI 0.51-2.24, P=0.86 [OxFU]) as compared to wild-types (OS; HR 1.04, 95% CI 0.87-1.23, P=0.67, P for interaction 0.99 and PFS; HR 0.96, 95% CI 0.82-1.12, P=0.60, P for interaction 0.40) and as compared to patients with other *KRAS* mutations (OS; HR 0.96, 95% CI 0.78-1.18, P=0.72, P for interaction 0.79 and PFS; HR 1.05, 95% CI 0.87-1.27, P=0.61, P for interaction 0.76).

Recent data (De Roock *et al.* 2010a) has highlighted how mutations of different exons of *PIK3CA* might alter response to cetuximab in different ways. Considering individual mutations from these exons (E542K, E545K and Q546K at exon 9 and H1047L and H1047R at exon 20), we show no evidence for differential effects on response.

### 3.3.6 Prognostic impact of somatic mutations

Irrespective of treatment received, median OS was shorter in those patients who had mutations in any of *KRAS, BRAF* and *NRAS* (n=706, 13.6 months) than among those whose tumours were all wild-type in these oncogenes (n=581, 20.1 months; log-rank test P<0.001, Figure 3.12). If the mutational type is separated, median OS is shorter in those patients who had mutations in *BRAF* (n=102, 8.8 months) than among those with a mutation in *KRAS* (n=548, 14.4 months) or *NRAS* (n=38, 13.8 months). A global test for differences in these curves was highly significant (P<0.001) as was a test for trend (P<0.001, Figure 3.13). A prognostic effect of patients' tumour mutation status was also observed on PFS, with median PFS time ranging from 5.6 months for patients who were *BRAF* mutant to 9.0 months for those patients with *KRAS, NRAS* and *BRAF* all wild-type.

We found no prognostic effects for *PIK3CA* mutations over PFS (any mutation [HR 1.06, 95% CI 0.89-1.26, P=0.50] or split between exons 9 [HR 1.09, 95% CI 0.89-1.33, P=0.43] and 20 [HR 0.98, 95% CI 0.73-1.33, P=0.91]).
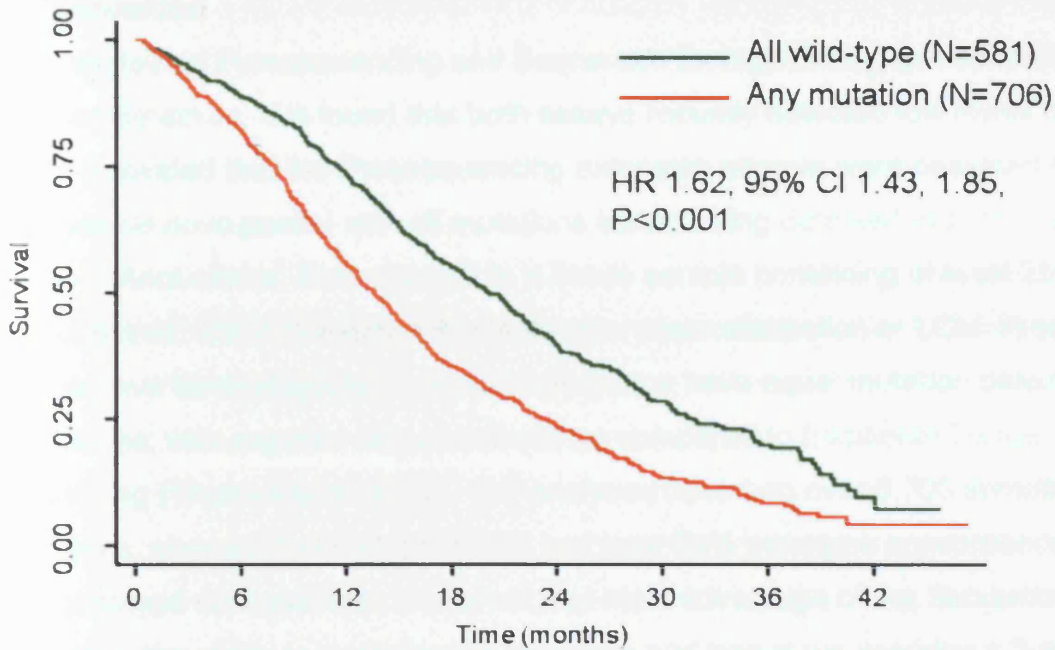
**Figure 3.12** – OS of patients with a mutation in any of *KRAS*, *BRAF* and *NRAS* versus those all wild-type, irrespective of treatment.
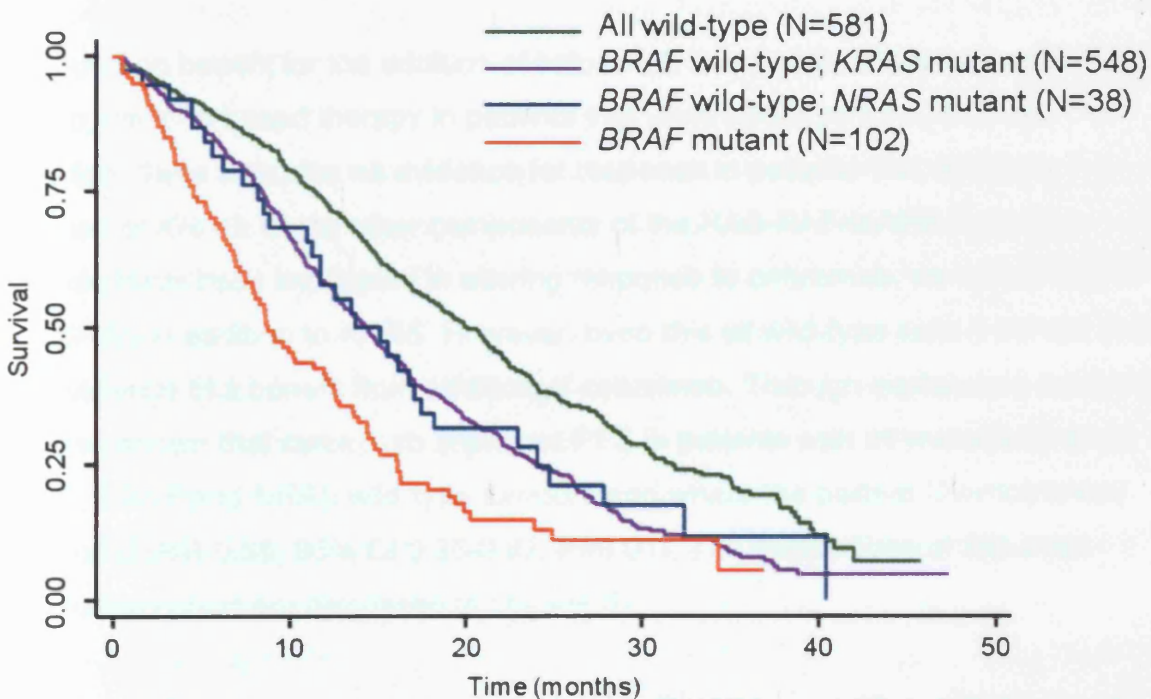


**Figure 3.13** – OS of patient all wild-type across *KRAS*, *BRAF* and *NRAS* (green) versus those wild-type for BRAF but with a mutation of *KRAS* (purple), those wild-type for *BRAF* but with a mutation of *NRAS* (blue) and those with only a mutation of *BRAF* (orange line).

## 3.4 Discussion

Here, we tested Pyrosequencing and Sequenom for high-throughput somatic mutation detection. We found that both assays robustly detected low levels of mutant alleles (provided that the Pyrosequencing extension primers were designed to generate *de novo* peaks) with all mutations tested being detected in samples with 12.5% mutant alleles. This equated to a tissue sample containing at least 25% tumour tissue, which is easily achievable after macrodissection or LCM. Previous studies have also suggested that these platforms have equal mutation detection sensitivities, with superior detection limits as compared to traditional Sanger sequencing (Thomas *et al.* 2007). Our analyses based on over 6,700 somatic genotypes, showed that both platforms had over 98% genotype concordance and high genotype success rates. However, the main advantage of the Sequenom platform is the ability to multiplex the reactions and herein we describe a 3-assay system to screen for a total of 33 somatic mutations within the Ras-Raf-MAP and PI3K-PTEN-Akt pathways. We propose that this validated system can now be introduced into routine clinical diagnostic practice.

We found no benefit for the addition of cetuximab to standard oxaliplatin and fluoropyrimidine based therapy in patients that were wild-type for *KRAS*. As expected, there was also no evidence for response in patients that carried a mutation of *KRAS*. Since other components of the RAS-RAF-MAPK signalling cascade have been implicated in altering response to cetuximab, we tested *BRAF* and *NRAS* in addition to *KRAS*. However, even this all wild-type cohort did not show any evidence of a benefit from addition of cetuximab. Through exploratory analysis, we have shown that cetuximab improved PFS in patients with ≤1 metastatic sites, *KRAS*, *BRAF* and *NRAS* wild-type tumours and where the partner chemotherapy was OxFU (HR 0.55, 95% CI 0.35-0.87, P=0.01). The implications of this most responsive cohort are discussed in chapter 07.

It has recently been reported that patients with *PIK3CA* exon 20 mutations treated with cetuximab have worse response rates, PFS and OS as compared with wild-type patients (DeRoock *et al.* 2010a). Therefore, we considered whether *PIK3CA* mutations might have a similar effect in our cohort. We did not find any detrimental effect for carriers of exon 20 mutations nor did we find that patients with *PIK3CA*

wild-type tumours (or those with exon 20 wild-type tumours) had improvements in OS or PFS upon treatment with cetuximab, regardless of chemotherapy regimen or somatic *KRAS* status. There was a significant association between *PIK3CA* exon 9 mutations and *KRAS*. The reason for this association may lie in the independent biological roles played by the different mutations of *PIK3CA*. Exon 9 mutations lie in the helical domain of the protein while exon 20 mutations lie in the kinase domain (Zhao *et al.* 2008). Though mutations of both exons result in activated Akt signalling, they have different requirements for interaction with the PI3K regulatory subunit p85 and with GTP bound Ras. Specifically, the gain of function coinciding with exon 9 mutations requires interaction with GTP bound Ras but is independent of binding to p85. Conversely, mutations of exon 20 require p85 binding but are independent of GTP bound Ras. As such, it has been suggested that mutations in these two domains represent two entirely distinct mutational states which differ in protein surface charge and conformational properties. This may explain the differences in cetuximab efficacy observed in the original De Roock (2010a) study, as well as the various other conflicting studies of *PIK3CA* (as discussed in chapter 1). In further support of there being distinct mutational states, the two exons have differing effects over breast cancer survival (Kalinsky *et al.* 2009; Lai *et al.* 2008).

It has also recently been reported that different somatic mutations within the Ras-Raf-MAP pathway have differential effects on response to cetuximab, with patients with *KRAS* p.G13D-mutated tumours benefitting from treated with cetuximab (De Roock *et al.* 2010b). However, we failed to find any such differences. Indeed, patients with G13D in *KRAS* treated with cetuximab trended towards worse OS and PFS, regardless of chemotherapy regimen or *KRAS* status. It may be that the observation of De Roock and colleagues is a consequence of the use of small cohorts of patients harbouring individual somatic mutations or a lack of control for confounding factors; of their pooled analysis of seven trials, only one (CO.17) randomised for cetuximab and in this group the effect of G13D was not significant.

Here we report that strongly significant prognostic roles were observed for mutant *KRAS* and *BRAF* with median survival ranging from 20.1 months for 'all wild-type' patients to 14 months for patients with a mutation of *KRAS* and 8.8 months for those with a *BRAF* mutation. The prognostic role of these genes (especially *KRAS*) had

been the subject of much debate prior to the completion of the COIN trial (as discussed in chapter 1). This data argues in the favour of these genes carrying important prognostic impact which should be considered in future trials as stratification factors or inclusion/exclusion criteria.

## 3.4.1 Future work

There is still much that is not understood about what causes patients to respond or not respond to certain treatments. The results of the COIN trial, combined with that of other clinical trials suggest that the correct clinical use of cetuximab in the treatment of CRC is still not entirely understood. Further studies will consider other components of the signalling pathways that operate downstream of EGFR such as PTEN, Akt, as well as those that operate in parallel pathways (e.g. IGF). Though not validated here, it is possible that individual mutations of these key genes could have different effects upon response. Larger cohorts with greater power will allow for a much greater understanding of the roles they play.

# Chapter Four - A variant in the *EIF3H* promoter influences colorectal cancer risk, patient survival and response to treatment

## 4.1 Introduction

Each year in the UK, more than 37,500 people are diagnosed with CRC (http://info.cancerresearchuk.org), with over a million new cases worldwide. The five-year relative survival rates for male and female CRC patients in the UK have doubled between the early 1970s and mid-2000s (from ~22-27% to ~50-55%) as a result of earlier diagnosis and better treatment (http://info.cancerresearchuk.org) (Rachet *et al.* 2009). Still there remains a need for improvements in the diagnosis and treatment of the disease. The search for risk factors (both high- and low-penetrance) of CRC is an ongoing effort that has provided some clues as to the aetiology of the disease.

High-throughput arrays for single nucleotide polymorphisms (SNPs) have opened new avenues for marker discovery by moving from hypothesis-driven targeted research to unbiased screens of the whole genome (Walther *et al.* 2009). To-date, these arrays have generally been used to search for disease susceptibility alleles by GWAS. Several studies have conducted multistage GWA for CRC and identified fourteen low penetrance susceptibility loci mapping across the genome. The role of these variants in aCRC with metastatic disease is unknown. Furthermore, none of these variants have been shown to affect patient outcome or response to treatment.

Here, we analysed twenty-two SNPs across ten of the fourteen loci identified from GWAS for CRC or that were highly significant in a subsequent meta-analysis, in a large collection of aCRC patients that had been recruited into the randomised controlled trials (RCTs) COIN and COIN-B, together with 2,176 healthy controls. We tested SNPs that contributed to the advanced setting for their prognostic potential and determined the underlying mechanism of action at 8q23.

## 4.2 Materials and Methods

### 4.2.1 Samples

We analysed 2,186 blood DNA samples from patients with aCRC from COIN (2,073 samples) and COIN-B (113 samples). Please refer to section 3.2.1 and Figure 3.1 for details of COIN samples. COIN-B is a MRC-funded, open-label, UK national phase II trial, examining intermittent chemotherapy plus cetuximab (target accrual of 158 patients). All patients gave fully informed consent for their samples to be used for bowel research. We also analysed 2,176 blood DNA samples from healthy controls from the UK Blood Services collection of Common Controls (UKBS collection). These samples were selected from a total of 3,092 samples within the UKBS collection that best matched the patients with aCRC in terms of place of residence within the UK. The UKBS samples had appropriate ethical approval. Clinicopathological features are described in Table 4.1.

### 4.2.2 Genotyping

Genotyping of twenty-three previously identified CRC risk alleles was performed by Illumina's Fast-Track Genotyping Services (San Diego, CA) using their high throughput BeadArray™ technology. Genotyping of rs28649280 was carried out by Geneservice (Nottingham, UK) using a TaqMan assay (Applied Biosystems) while rs16888589 was genotyped by Kbiosciences (Herts, UK).

|  |  | COIN (%) | COIN-B (%) | Total cases (%) | Controls (%) |
|---|---|---|---|---|---|
| n = |  | 2073 | 113 | 2186 | 2176 |
| Age at diagnosis (aCRC)/Sampling (controls) | Mean | 61.54 | 61.16 | 61.51 | 43.7 |
|  | <20 | 1 (0.0) | 0 | 1 (0.0) | 64 (2.9) |
|  | 20-49 | 232 (11.2) | 13 (11.5) | 245 (11.2) | 1317 (60.2) |
|  | 50-59 | 549 (26.5) | 27 (23.9) | 576 (26.3) | 602 (27.7) |
|  | 60-69 | 845 (40.8) | 49 (43.4) | 894 (40.9) | 193 (8.9) |
|  | 70-79 | 435 (21.0) | 22 (19.5) | 457 (20.9) | 0 |
|  | 80-89 | 9 (0.1) | 2 (1.8) | 11 (0.5) | 0 |
|  | Missing | 2 (0.1) | 0 | 2 (0.1) | 0 |
| Sex | Female | 698 (33.7) | 48 (42.5) | 746 (34.1) | 1074 (49.4) |
|  | Male | 1375 (66.3) | 65 (57.5) | 1440 (65.9) | 1102 (50.6) |
| WHO-PS | 0 | 969 (46.7) | 58 (51.3) | 1027 (47.0) | - |
|  | 1 | 951 (45.9) | 46 (40.7) | 997 (45.6) | - |
|  | 2 | 153 (7.4) | 9 (8.0) | 162 (7.4) | - |
| Primary Site | Colon | 1119 (54.0) | 37 (32.7) | 1156 (52.9) | - |
|  | Rectum | 653 (31.5) | 32 (28.3) | 685 (31.3) | - |
|  | RSJ | 297 (14.3) | 10 (8.8) | 307 (14.0) | - |
|  | Other | 3 (0.1) | 34 (30.1) | 37 (1.7) | - |
|  | Missing | 1 (0.0) | 0 | 1 (0.0) | - |
| Number of metastatic sites | 0 | 14 (0.7) | 1 (0.9) | 15 (0.7) | - |
|  | 1 | 737 (35.6) | 43 (38.1) | 780 (35.7) | - |
|  | 2 | 815 (39.3) | 50 (44.2) | 865 (39.6) | - |
|  | ≥3 | 507 (24.5) | 19 (16.8) | 526 (24.1) | - |
| Metastatic sites | Liver only | 459 | 24 | 483 | - |
|  | Liver + Nodal | 691 | 34 | 725 | - |
|  | Liver + Lungs | 594 | 24 | 618 | - |
|  | Liver + Peritoneum | 202 | 9 | 211 | - |
|  | Liver + Other | 184 | 15 | 199 | - |
|  | No Liver | 520 | 33 | 553 | - |

**Table 4.1** - Clinicopathological data for cases (COIN and COIN-B) and controls. Abbreviations: RSJ - Rectosigmoid Junction

101

## 4.2.3 RT-PCR and RACE

RT-PCR was carried out using human colonic first strand cDNA (Stratagene) with forward primers in either the first (RT1F 5'-GCACAAAGGATACCGCCTGGAAAG-3') or second (RT2F 5'-CTTCCTGTCTGCTTGGAAAGATGG-3') predicted *EIF3H* exons in combination with reverse primers in exons 2 (RT2R 5'-GGCCATCTATCTGCACTTGCTTCAC-3'), 3 (RT3R 5'-CAAAGTCAGCATCATCCTCTGTGTGC-3'), 4 (RT4R 5'-AGGAGTGCCCGGGTAACGAATGAG-3') and 5 (RT5R 5'-CCTTTAGTGAGAGAGATCCTTGGGCAGT-3'). Products were analysed on 2% agarose gels. RACE was performed on human colonic Marathon ready cDNA using the primers (RT2R) and (RT4R). Products were cloned into pGEM-T Easy, transformed into JM109 *E.coli* and sequenced.

## 4.2.4 DNA and RNA extraction from FFPE CRCs and MSI analyses

10μm sections were cut from FFPE CRCs and stained with haematoxylin and eosin. Tumour material was obtained by macrodissection or laser capture microdissection. DNA was extracted using DNA microkits and analysed for MSI using the markers BAT-25 and BAT-26. RNA was extracted using the 'RecoverAll Total Nucleic Acid Isolation' Kit (Ambion) and converted to cDNA using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems).

## 4.2.5 Real-Time PCR

EIF3H targeted Taqman assays Hs01052033_m1 (spanning exons 2 and 3), and Hs00186779_m1 (spanning exon 6 and 8) (Applied Biosystems) were used in gene expression analysis, along with Taqman Universal PCR master mix. Human B2M was used as an endogenous control (Kheirelseid *et al.* 2010).

## 4.2.6 Immunohistochemistry

Immunohistochemistry was carried out using primary antibodies against EIF3H (D9C1, Cell signalling), FGF-2 (sc-79; Santa Cruz Biotechnology), C-myc (sc-40; Santa Cruz Biotechnology) and cyclin-D1 (sc-8396; Santa Cruz Biotechnology). Staining was scored blind with respect to the genotypes.

### 4.2.7 Luciferase assay

The potential EIF3H promoter region was PCR amplified using the primers; (LUC ASSAY F4 – 5'–ACTAAGCTTATACCGCCTGGAAAGAAGGT-3'; LUC ASSAY R4 - 5'- TACTGCCATGGCCAAGCAGACAGGAAGAAAGA-3' and ligated into a reporter construct which was then transformed into JM109 *E.coli*. The luciferase vector and internal control plasmid DNA (Promega's pRL-CMV), were transiently transfected into Human Embryonic Kidney 293 cells. Six replicates were transfected with each reporter construct and each transfection experiment was repeated at least four times. The relative luciferase activity was determined by comparing luciferase activity of pGL3 constructs against a pRL control reporter.

### 4.2.8 Statistical analyses

Single marker, haplotype and interaction association analyses were performed using PLINK software (Purcell *et al.* 2007). To correct for multiple testing, we carried out 10,000 permutations taking into account the LD between the markers. Case-only analysis of the clinicopathological data was performed using step-wise regression analyses using SPSS statistical software. Gender, WHO performance status (PS), number of metastatic sites, liver only metastases and MSI were analysed using logistic regression (coded as 0/1), age at diagnosis was analysed using linear regression and the primary site was analysed using nominal regression. This analysis was performed by Dr. Valentina Moskvina.

A Cox survival model was used, with OS as the outcome. The model-building process was: (i) each SNP was entered into a univariate model to test for individual prognostic value, (ii) all SNPs were entered into a single multivariate model to test for prognostic value independent of one-another, (iii) all SNPs were added to a model containing the known prognostic factors (WHO PS, number of metastatic sites, white blood cell [WBC] count, alkaline phosphatase levels and *KRAS* and *BRAF* mutation status) to determine whether they had additional prognostic value, (iv) all SNPs and known prognostic factors were entered into a closed-test procedure multiple fractional polynomial (FP) model in order to construct the most efficient model (a *P*-value of 0.05 was used for all covariates for selection in the model).

Prognostic analysis of toxicity (Arms A and C) – gastrointestinal toxicity was defined as a per-patient maximum recorded CTC grade 3-5 for nausea, vomiting or diarrhoea and compared against a per-patient maximum CTC grade of 0-1. Haematological toxicity was defined as a per-patient maximum recorded CTC grade 3-5 for platelets, haemoglobin, WBC or neutrophils and compared to a per-patient maximum CTC grade of 0-1. Skin toxicity was defined as a per-patient maximum recorded CTC grade 3-5 for nail changes, skin rash or hand-foot syndrome and compared to a per-patient maximum CTC grade of 0-1. For all toxicities, patients with a CTC grade of 2 were excluded in order to have a stronger differentiation between the most and least severe groups. All survival and toxicity based statistical analyses were run by David Fisher of the MRC CTU.

Statistical analyses for the real-time RT-PCR assays was performed using Bland-Altman (Bland and Altman, 1986) and two-sample T tests, and for the immunohistochemical staining using Mantel-Haenszel Chi-Squared tests. Differences in relative luciferase activity were assessed using the Mann-Whitney test.

## 4.3 Results

### 4.3.1 Genotyping

Twenty-four SNPs that had been identified from GWAS as being associated with CRC (Tomlinson *et al.* 2007; Broderick *et al.* 2007; Jaeger *et al.* 2008; Tenesa *et al.* 2008; Tomlinson *et al.* 2008; Houlston *et al.* 2008) or that were significant at $P<10^{-5}$ in a subsequent meta analysis (Houlston *et al.* 2008), were analysed with Illumina's Assay Design Tool to determine their potential for locus conversion on the GoldenGate platform. One SNP (rs4951039) failed *in silico* locus conversion. The remaining twenty-three SNPs were genotyped in our case-control series (Figure 4.1). rs7837328 failed genotyping. For the remaining 22 SNPs, genotyping concordance rates for duplicate samples was 100% (1022/1022 genotypes were concordant), GenTrain scores ranged from 0.665 to 0.9655 and the overall genotyping success rate was 99% (95027/95964 genotypes were called successfully).

### 4.3.2 Associations with aCRC

We found that eleven SNPs representing five different loci were significantly over-represented in aCRC cases or healthy controls at the 5% level (Table 4.2). After adjustment for multiple testing, rs16892766 at 8q23, rs10808555 and rs7014346 at 8q24, rs4779584 at 15q13, and rs4939827 and rs12953717 at 18q21 remained significantly over-represented (Table 4.2). For each of these loci at least two SNPs were genotyped, therefore we tested whether combinations of SNPs at each locus altered risk. Using step-wise logistic regression analysis on the basis of an additive model, we found that only a single SNP at each locus was responsible for the disease association (rs4939827 at 18q21, rs16892766 at 8q23, rs4779584 at 15q13 and rs10808555 at 8q24) with the other SNPs being in LD and moderately correlated with these risk alleles (rs12953717 tagged rs4939827 [$r^2$=0.623, D'=0.931], rs11986063 and rs6983626 tagged rs16892766 [$r^2$=0.427 and 0.746, D'=0.720 and 0.932, respectively], rs10318 tagged rs4779584 [$r^2$=0.550, D'=0.748], and rs6983267 and rs7014346 tagged rs10808555 [$r^2$=0.312 and 0.608, D'=0.824 and 0.840, respectively]).
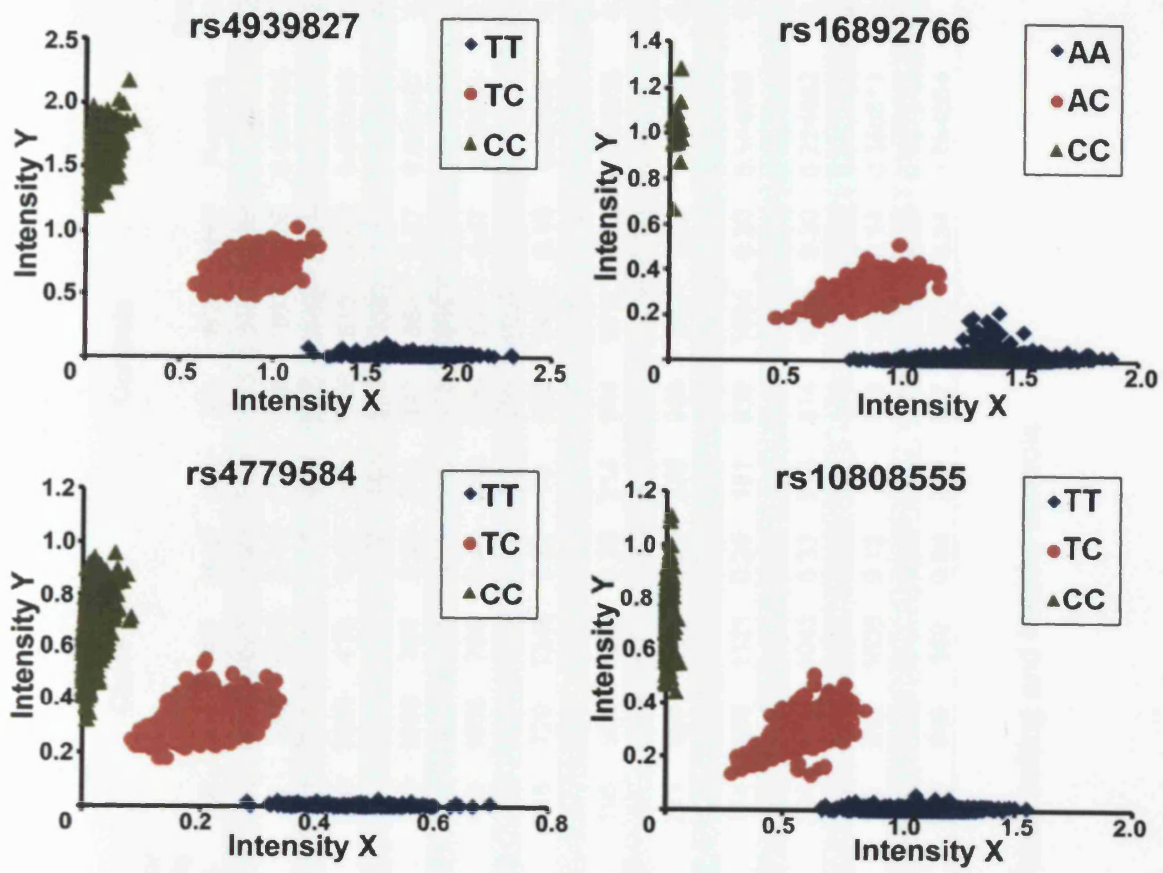
**Figure 4.1** - Cluster plots of rs4939827 (18q21), rs16892766 (8q23), rs4779584 (15q13), and rs10808555 (8q24), generated by Illumina's GoldenGate assay.

| SNP number | Chromosome | Position (build 37) | GenTrain Score | OR | Minor allele (A) | Major allele (B) | Cases AA | AB | BB | MAF | Controls AA | AB | BB | MAF | P-value | Corrected P-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs4939827 | 18q21 | 46453463 | 0.8975 | 0.85 | C | T | 433 | 1082 | 666 | 0.45 | 504 | 1113 | 558 | 0.49 | 0.000114 | 0.0027 |
| rs16892766 | 8q23 | 117630683 | 0.7024 | 1.33 | C | A | 15 | 396 | 1772 | 0.10 | 14 | 299 | 1862 | 0.08 | 0.000183 | 0.0034 |
| rs4779584 | 15q13 | 32994756 | 0.7922 | 1.21 | T | C | 123 | 737 | 1323 | 0.23 | 81 | 677 | 1413 | 0.19 | 0.000286 | 0.0059 |
| rs12953717 | 18q21 | 46453929 | 0.665 | 1.15 | T | C | 359 | 1064 | 419 | 0.48 | 311 | 1086 | 516 | 0.45 | 0.000439 | 0.0097 |
| rs10808555 | 8q24 | 128409511 | 0.8554 | 1.15 | C | T | 313 | 992 | 877 | 0.37 | 257 | 959 | 959 | 0.34 | 0.001923 | 0.0378 |
| rs7014346 | 8q24 | 128424792 | 0.8166 | 1.14 | A | G | 369 | 1030 | 784 | 0.40 | 314 | 993 | 866 | 0.37 | 0.002487 | 0.0477 |
| rs10795668 | 10p14 | 8701219 | 0.7972 | 0.87 | A | G | 196 | 909 | 986 | 0.31 | 226 | 968 | 889 | 0.34 | 0.003336 | 0.0635 |
| rs6983267 | 8q24 | 128413305 | 0.7725 | 0.89 | T | G | 439 | 1035 | 704 | 0.44 | 483 | 1067 | 617 | 0.47 | 0.00565 | 0.1029 |
| rs11986063 | 8q23 | 117640315 | 0.9655 | 1.22 | T | C | 22 | 427 | 1734 | 0.11 | 20 | 352 | 1801 | 0.09 | 0.005696 | 0.1034 |
| rs10318 | 15q13 | 33025979 | 0.8188 | 1.16 | T | C | 118 | 720 | 1345 | 0.22 | 78 | 692 | 1405 | 0.19 | 0.00596 | 0.1092 |
| rs6983626 | 8q23 | 117802148 | 0.8584 | 1.18 | T | C | 27 | 429 | 1727 | 0.11 | 19 | 376 | 1780 | 0.10 | 0.017557 | 0.2788 |
| rs1862748 | 16q22 | 68832943 | 0.8914 | 0.92 | T | C | 196 | 904 | 1083 | 0.30 | 214 | 944 | 1016 | 0.32 | 0.058368 | 0.6613 |
| rs7259371 | 19q13 | 33534641 | 0.9362 | 1.11 | A | G | 89 | 660 | 1431 | 0.19 | 82 | 602 | 1483 | 0.18 | 0.067813 | 0.7128 |
| rs355527 | 20p12 | 6388068 | 0.7663 | 1.08 | A | G | 261 | 990 | 932 | 0.35 | 236 | 956 | 983 | 0.33 | 0.07512 | 0.7486 |
| rs961253 | 20p12 | 6404281 | 0.8814 | 1.07 | A | C | 305 | 1034 | 843 | 0.38 | 281 | 1008 | 886 | 0.36 | 0.125169 | 0.8985 |
| rs9929218 | 16q22 | 68820946 | 0.9119 | 0.93 | A | G | 174 | 886 | 1121 | 0.28 | 181 | 930 | 1064 | 0.30 | 0.144065 | 0.9279 |
| rs4444235 | 14q22 | 54410919 | 0.6814 | 1.06 | C | T | 513 | 1059 | 607 | 0.48 | 459 | 1105 | 611 | 0.47 | 0.211965 | 0.9799 |
| rs3802842 | 11q23 | 111171709 | 0.8738 | 1.06 | C | A | 224 | 916 | 1043 | 0.31 | 216 | 874 | 1085 | 0.30 | 0.224882 | 0.9852 |
| rs10411210 | 19q13 | 33532300 | 0.8834 | 1.07 | T | C | 25 | 380 | 1778 | 0.10 | 23 | 356 | 1795 | 0.09 | 0.343304 | 0.9988 |
| rs4951291 | 1q32 | 204006538 | 0.7298 | 0.95 | T | C | 29 | 512 | 1636 | 0.13 | 29 | 538 | 1600 | 0.14 | 0.356811 | 0.9993 |
| rs10749971 | 11q23 | 111189158 | 0.8868 | 1.02 | G | A | 316 | 1018 | 847 | 0.38 | 309 | 1008 | 858 | 0.37 | 0.667828 | 1 |
| rs11213809 | 11q23 | 111135745 | 0.7223 | 1.01 | A | G | 263 | 949 | 969 | 0.34 | 269 | 927 | 979 | 0.34 | 0.894664 | 1 |

**Table 4.2** – GWAS SNPs for CRC analysed in the advanced disease setting and healthy controls.

For rs4939827 (OR 0.85, 95% CI 0.78-0.93), the minor allele was associated with a decreased risk of CRC in a dose-dependent manner (OR$_{het}$ 0.88, 95% CI 0.76-1.03, OR$_{hom}$ 0.72, 95% CI 0.61-0.85). For rs16892766 (OR 1.33, 95% CI 1.14-1.55), genotype-specific ORs were most compatible with an over-dominant model (OR$_{ACvs.AA}$ 1.39, 95% CI 1.18-1.64; OR$_{ACvs.CC}$ 1.24, 95% CI 0.59-2.60). For rs4779584 (OR 1.21, 95% CI 1.09-1.35), the minor allele was associated with an increased risk of CRC in a dose-dependent manner (OR$_{het}$ 1.39, 95% CI 1.03-1.88; OR$_{hom}$ 1.62, 95% CI 1.21-2.17). A similar risk pattern was also observed for rs10808555 (OR 1.15, 95% CI 1.05-1.26; OR$_{het}$ 1.18, 95% CI 0.98-1.42; OR$_{hom}$ 1.33, 95% CI 1.10-1.61). We tested for potential haplotype effects for all markers at each of the significant loci. Although we found potential haplotype effects at the 8q24 and 18q21 loci, the changes in risk were small and insignificant after correction for multiple testing.

Using case-only logistic regression analysis, we tested for associations between clinicopathological variables such as gender, age at diagnosis, WHO PS, primary site, number and location of metastatic sites and MSI, and the genotypes of rs4939827, rs16892766, rs4779584 and rs10808555. Although a number of associations were observed including rs4779584 and age at diagnosis and WHO PS, rs9929218 and number of metastatic sites, rs961253 and liver metastasis and rs7259371 and location of the primary tumour, none remained significant after correction for multiple testing.

### 4.3.3 Associations with patient survival

We tested whether rs4939827, rs16892766, rs4779584 or rs10808555 influenced patient survival by utilising clinical data from COIN. We found no imbalances in the frequencies of the four SNPs tested between the three treatment arms. We have previously shown that *KRAS*, *BRAF* and *NRAS* mutations are found in 42.27%, 9.01% and 3.56% of the CRCs from patients in COIN (all trial Arms; Maughan *et al.* 2011). We found no differences between the frequencies or the genotype distributions of the SNPs according to *KRAS*, *BRAF* and *NRAS* mutation status (Table 4.3).

| | KRAS wild-type N (%) | mutant N (%) | BRAF wild-type N (%) | mutant N (%) | NRAS wild-type N (%) | Mutant N (%) |
|---|---|---|---|---|---|---|
| **rs4939827** | | | | | | |
| TT | 311 (31%) | 202 (28%) | 469 (30%) | 43 (28%) | 498 (30%) | 16 (26%) |
| TC | 480 (49%) | 372 (52%) | 774 (50%) | 77 (51%) | 813 (50%) | 32 (52%) |
| CC | 198 (20%) | 142 (20%) | 307 (20%) | 32 (21%) | 326 (20%) | 13 (21%) |
| Total | 989 (100%) | 716 (100%) | 1,550 (100%) | 152 (100%) | 1,637 (100%) | 61 (100%) |
| **rs16892766** | | | | | | |
| AA | 813 (82%) | 564 (79%) | 1,241 (80%) | 134 (88%) | 1,323 (81%) | 51 (84%) |
| AC | 170 (17%) | 148 (21%) | 299 (19%) | 17 (11%) | 303 (19%) | 10 (16%) |
| CC | 7 (1%) | 4 (1%) | 11 (1%) | 1 (1%) | 12 (1%) | 0 (0%) |
| Total | 990 (100%) | 716 (100%) | 1,551 (100%) | 152 (100%) | 1,638 (100%) | 61 (100%) |
| **rs4779584** | | | | | | |
| CC | 603 (61%) | 416 (58%) | 927 (60%) | 91 (60%) | 977 (60%) | 37 (61%) |
| CT | 340 (34%) | 249 (35%) | 533 (34%) | 55 (36%) | 566 (35%) | 22 (36%) |
| TT | 47 (5%) | 51 (7%) | 91 (6%) | 6 (4%) | 95 (6%) | 2 (3%) |
| Total | 990 (100%) | 716 (100%) | 1,551 (100%) | 152 (100%) | 1,638 (100%) | 61 (100%) |
| **rs10808555** | | | | | | |
| TT | 400 (40%) | 292 (41%) | 636 (41%) | 55 (36%) | 655 (40%) | 33 (54%) |
| TC | 447 (45%) | 320 (45%) | 687 (44%) | 78 (51%) | 747 (46%) | 18 (30%) |
| CC | 143 (14%) | 104 (15%) | 228 (15%) | 19 (13%) | 236 (14%) | 10 (16%) |
| Total | 990 (100%) | 716 (100%) | 1,551 (100%) | 152 (100%) | 1,638 (100%) | 61 (100%) |

**Table 4.3** - Genotype distributions of rs4939827, rs16892766, rs4779584 and rs10808555, and *KRAS*, *BRAF* and *NRAS* mutation status

We tested for, but did not find, any effect for rs4939827, rs4779584 or rs10808555 on OS (Figure 4.2, Table 4.4). However, for rs16892766, we found that patients carrying one or more of the minor alleles (AC or CC) showed a significant decline in OS as compared to patients carrying the AA genotype (HR 1.27, 95% CI 1.12-1.44, $P<0.001$), equating to a median decrease in life expectancy of 2.9 months (Figure 4.2, Table 4.4).

To determine whether this finding was influenced by the type and duration of treatment, we evaluated rs16892766 according to trial Arm. The effects were significant across Arms A and B (HR 1.32, 95% CI 1.06-1.64, P=0.012 and HR 1.45, 95% CI 1.17-1.79, P=0.001 respectively); however for patients treated with intermittent oxaliplatin and fluoropyrimidine chemotherapy (Arm C), the effect did not reach statistical significance (HR 1.10, 95% CI 0.89-1.36, P=0.367), suggesting that rs16892766 may influence survival based upon an interaction with treatment.

We have previously shown that the WHO PS, number of metastatic sites, WBC count, alkaline phosphatase levels and *KRAS* and *BRAF* mutation status are independent prognostic factors affecting OS in patients with aCRC. We tested a multivariate model including these known prognostic factors and found that patients carrying one or more of the minor alleles for rs16892766 showed a significant decline in OS, independent of the known prognostic factors (Table 4.5).

### 4.3.4 Prognostic effects on response to and side effects from 12 weeks of chemotherapy

We tested to see if any of rs4939827, rs16892766, rs4779584 and rs10808555 altered response to and side-effects from 12 weeks of oxaliplatin and fluoropyrimidine therapy (COIN Arms A and C only). We found no significant observations. Similarly, we tested these SNPs for a predictive effect on response to and side effects from 12 weeks of cetuximab treatment (COIN Arm B vs. Arms A and C combined) but found no significant observations.

### 4.3.5 Delineating the mechanism underlying rs16892766

rs16892766 at 8q23.3 lies 26.4kb downstream of *EIF3H*, whereas the weaker CRC-associated SNPs rs6983626 and rs11986063 lie 34.1kb upstream and 16.7kb downstream, respectively (Figure 4.3). Apart from a predicted transcript (*UTP23/C8orf53*) no other genes lie within this region, suggesting that *EIF3H* may play a direct role in colorectal tumourigenesis and survival outcome.
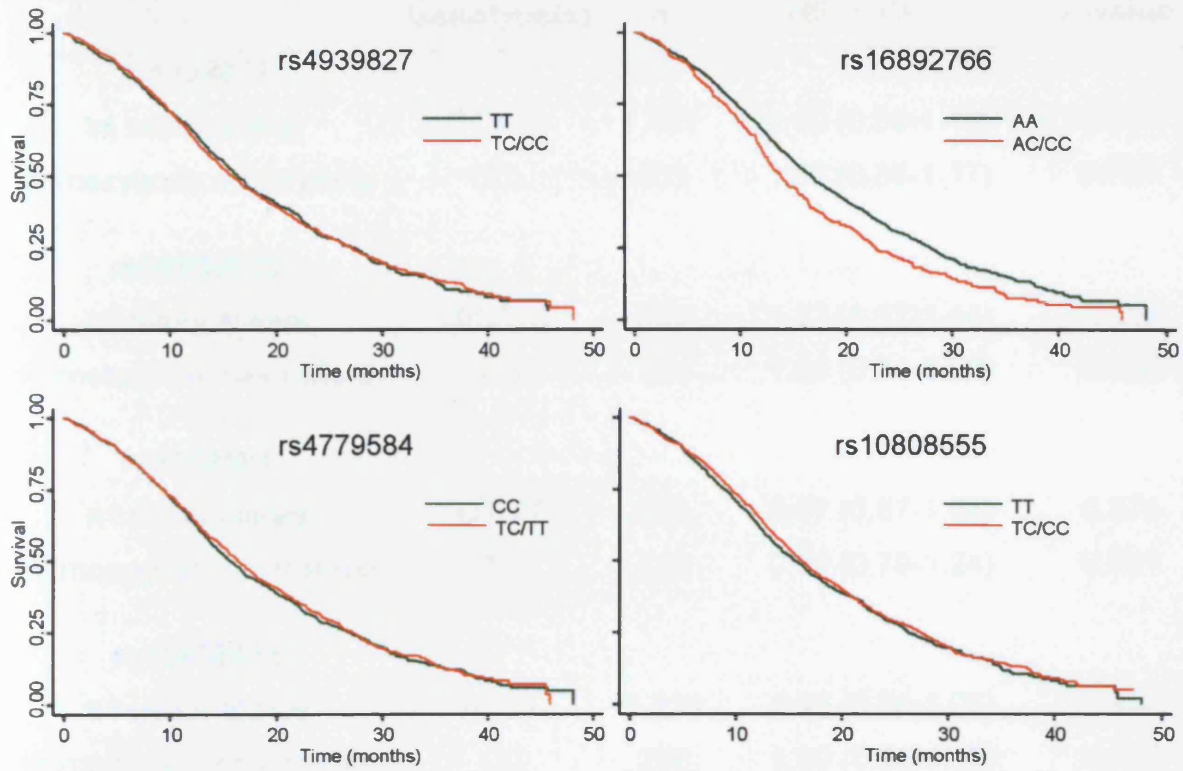
**Figure 4.2** - Kaplan-Meier curves for the effect of rs4939827, rs16892766, rs4779584, and rs10808555, on OS.

| | Genotype(s) | n | HR (95% CI) | P-value |
|---|---|---|---|---|
| **rs4939827** | | | | |
| ≥1 minor alleles | TC/CC | 1,431 | 1.00 (0.89-1.12) | 0.954 |
| Homozygous minor allele | CC | 403 | 1.02 (0.89-1.17) | 0.761 |
| | | | | |
| **rs16892766** | | | | |
| ≥1 minor alleles | AC/CC | 389 | 1.27 (1.12-1.44) | <0.001 |
| Homozygous minor allele | CC | 13 | 1.00 (0.54-1.89) | 0.989 |
| | | | | |
| **rs4779584** | | | | |
| ≥1 minor alleles | CT/TT | 825 | 0.97 (0.87-1.08) | 0.574 |
| Homozygous minor allele | TT | 115 | 0.99 (0.79-1.24) | 0.931 |
| | | | | |
| **rs10808555** | | | | |
| ≥1 minor alleles | TC/CC | 1,234 | 0.96 (0.86-1.07) | 0.422 |
| Homozygous minor allele | CC | 298 | 1.00 (0.86-1.16) | 0.972 |

**Table 4.4** – Univariate analyses use Cox proportional-hazards models with the outcome of OS. Hazard ratios should be interpreted as the hazard in the genotype group indicated versus all other patients with a valid test result for that SNP.

| Variable | | HR (95% CI) | P-value |
|---|---|---|---|
| rs16892766 | AC and CC vs. AA | 1.28 (1.13-1.45) | <0.001 |
| | CC vs. AC and AA | 1.24 (0.64-2.42) | 0.528 |
| WHO PS | 1 vs. 0 | 1.26 (1.12-1.41) | <0.001 |
| | 2 vs. 0 | 1.73 (1.39-2.16) | <0.001 |
| No. of metastatic sites | ≥2 vs. 0 or 1 | 1.26 (1.12-1.42) | <0.001 |
| WBC | ≥10,000/μl vs. <10,000/μl | 1.50 (1.33-1.71) | <0.001 |
| Alkaline phosphatase | ≥300U/l vs. <300U/l | 1.71 (1.47-1.99) | <0.001 |
| KRAS status | Mutant vs. wild-type | 1.37 (1.12-1.55) | <0.001 |
| BRAF status | Mutant vs. wild-type | 2.34 (1.93-2.84) | <0.001 |

**Table 4.5** - Multivariate model including known prognostic factors. Table shows the results from a single, multivariate, Cox model with the outcome of OS (n=1,704 patients).
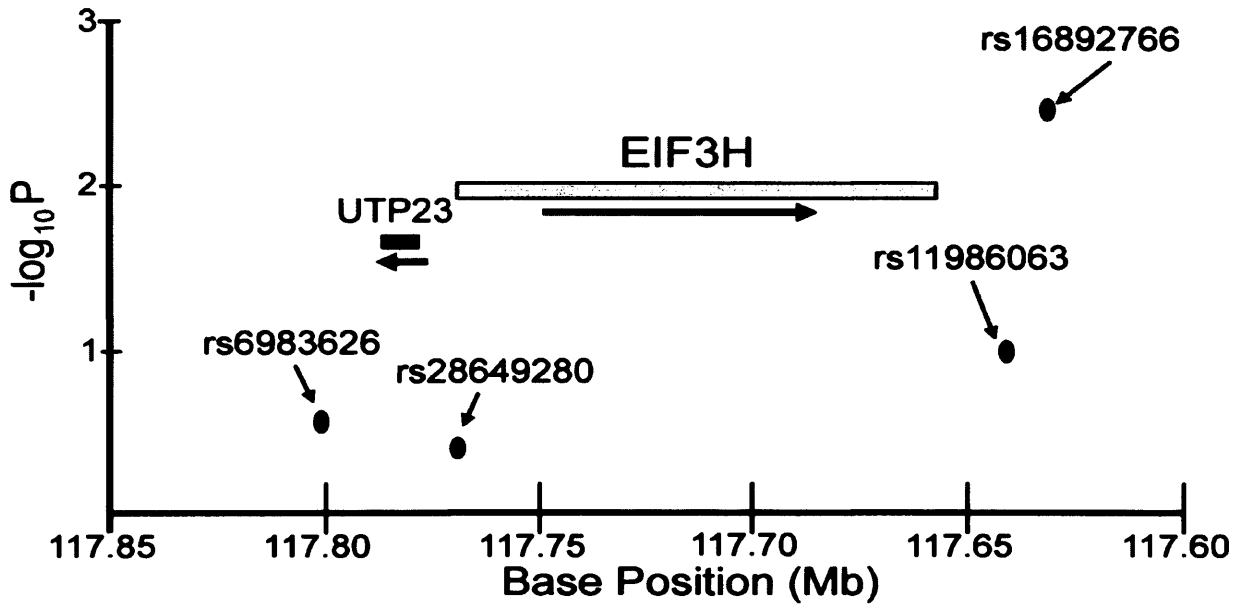
**Figure 4.3** - Regional plot of the 8q23 region. Shown are the association statistics (as -log$_{10}$P) together with relative positions of *EIF3H* and *UTP23*.

### 4.3.6 *EIF3H* sequencing

To identify potential pathogenic variants within *EIF3H* that were linked to rs16892766, we sequenced the entire *EIF3H* open reading frame and flanking intronic sequences in 124 aCRC cases (15 with the rs16892766-CC genotype, 94 with the AC genotype and 15 with the AA genotype) and 126 healthy controls (14 with the CC genotype, 97 with the AC genotype and 15 with the AA genotype). By using a total of 250 samples, we had over 90% power (0.05 significance level) to detect rare variants with MAFs of 0.005. *In silico* analyses of the *EIF3H* locus predicted two main transcripts which differed in their 5' sequences; the largest transcript EIF3H-202 spanned 9 exons and contained an additional upstream exon as compared to EIF3H-201 (Figure 4.4). In total, we identified a single variant (IVS1-24 C>T) in one control sample and a common polymorphism (rs28649280) in the first predicted intron.

### 4.3.7 Analysis of rs28649280 in aCRC cases and controls

We genotyped rs28649280 in 2,186 cases with aCRC and 2,176 UKBS healthy controls (Figure 4.5). The genotype success rate was 96.5% and the concordance rate for duplicated samples was 100%. We found that rs28649280 conformed to the Hardy Weinberg equilibrium ($P$=0.867) and the minor allele was significantly over-represented in cases (MAF=0.11) as compared to controls (MAF=0.095, $P$=0.022). Genotype-specific ORs (OR 1.18, 95% CI 1.025-1.359) were most compatible with a multiplicative model of disease susceptibility ($OR_{het}$ 1.21, 95% CI 0.66-2.21; $OR_{hom}$ 1.42, 95% CI 0.78-2.57). We observed that rs28649280 was in moderate LD with rs16892766 ($r^2$=0.441, D'=0.734) and a logistic regression model for increased risk with both SNPs combined was not more significant as compared to rs16892766 on its own ($P$=0.84). We did not observe any associations between the clinicopathological variables and rs28649280 genotype.
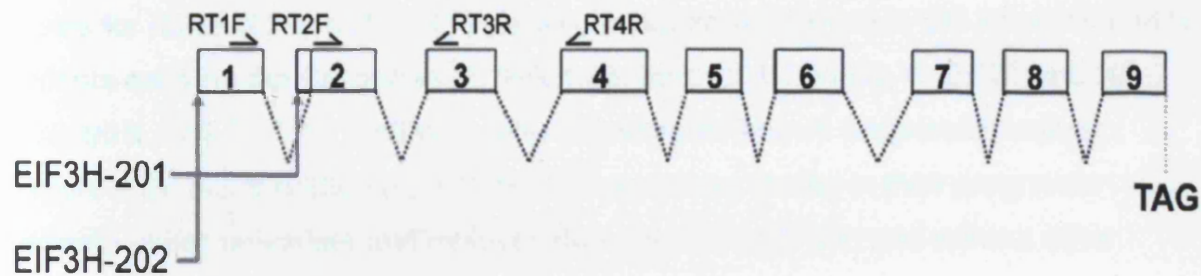
**Figure 4.4** - Diagrammatic representation of the predicted *EIF3H* transcripts (boxes represent exons and dotted lines represent splicing patterns). EIF3H-201 starts at exon 2, whilst EIF3H-202 starts at exon 1 and splices within exon 2. The positions of the primers used for RT-PCR and RACE are indicated.
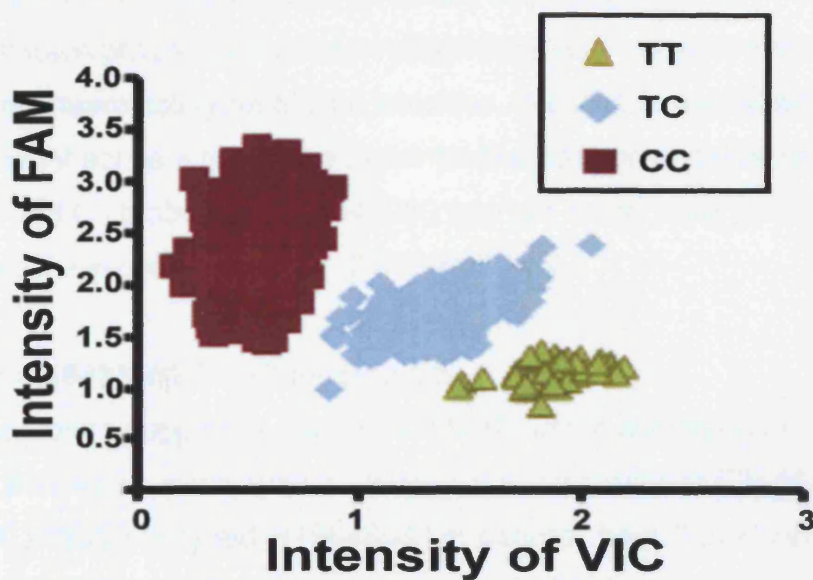


**Figure 4.5** - Cluster plot defining the rs28649280 genotype of 2,186 aCRC cases and 2,176 UKBS healthy controls.

### 4.3.8 rs28649280 influences OS and response to treatment

We found that under a univariate analysis, patients carrying one or more of the minor alleles for rs28649280 (CT or TT) showed a significant decline in OS as compared to patients carrying the CC genotype (HR 1.26, 95% CI 1.12-1.43, $P<0.001$ and HR 1.30, 95% CI 1.13-1.49, $P<0.001$, after adjusting for known prognostic factors) (Figure 4.6). rs28649280 and rs16892766 were very similar in their prognostic strength under univariate and multivariate analyses both with and without other known prognostic variables.

We also examined rs28649280 for its potential prognostic effect on response to 12 weeks of treatment with oxaliplatin and fluoropyrimidine (COIN Arms A and C only). We found that patients carrying one or more of the minor alleles showed a significant decline in 12 week response as compared to patients carrying the CC genotype (OR 0.65, 95% CI 0.48-0.87, $P=0.004$ and OR 0.60, 95% CI 0.43-0.83, $P=0.002$, after adjusting for known prognostic factors). rs28649280 was not associated with gastrointestinal, haematological or skin toxicities. We also tested whether rs28649280 might act as a predictive factor for the addition of cetuximab (COIN Arm B vs. Arms A and C combined). rs28649280 was not found to be predictive factor in terms of 12-week response ($\chi^2$=3.04, 2 d.f., $P=0.218$).

### 4.3.9 Role of rs16888589 in influencing OS

Recently, it has been suggested that rs16888589, which is in strong LD with rs16892766, acts as an allele specific, long-range, repressor of EIF3H transcription (Pittman et al. 2010). We typed rs16888589 in patients from COIN and COIN-B and found that it had similar prognostic effects to rs28649280 and rs16892766 on OS (HR 1.21, 95% CI 1.07–1.37, $P=0.002$) and 12-week response to treatment (OR 0.72, 95% CI 0.54–0.97, $P=0.031$).

### 4.3.10 RT-PCR and RACE

We investigated the mechanism of action of rs28649280. First, we clarified the structure of *EIF3H* to determine whether the 1st predicted exon was transcribed within the colon. We carried out RT-PCR using cDNA from healthy colonic tissue using primers spanning exons 1-5 of *EIF3H* (Figure 4.4). We failed to generate any product using primers within the first predicted exon, but did generate products from

the downstream exons (Figure 4.7, left). We also carried out 5'RACE on normal colonic cDNA, cloned and sequenced the products (Figure 4.7, right). These results confirmed that the first predicted exon was not transcribed within the colon and that the transcription start site was 14bp upstream from the predicted kozak translation initiation site in EIF3H-201 (Figure 4.8). rs28649280 therefore lies within a potential promoter region, 78bp upstream of the transcription start site used within the colon.

### 4.3.11 *In silico* analysis of the *EIF3H* promoter

We analysed the potential *EIF3H* promoter region used in the colon for transcription factor binding sites using TFSEARCH and determined their conservation through evolution by analysis of human, chimpanzee, dog, rhesus macaque, mouse and rat sequences using ClustalW. rs28649280 was found to lie in a potential Sp1 binding site (CCGCCCY) and the whole 7bp sequence was fully conserved in all species examined (Figure 4.8). Other conserved basal transcription factor binding sites (including CREB and Ets) were found adjacent to the Sp1 site, in support of this region acting as a promoter (Figure 4.8).
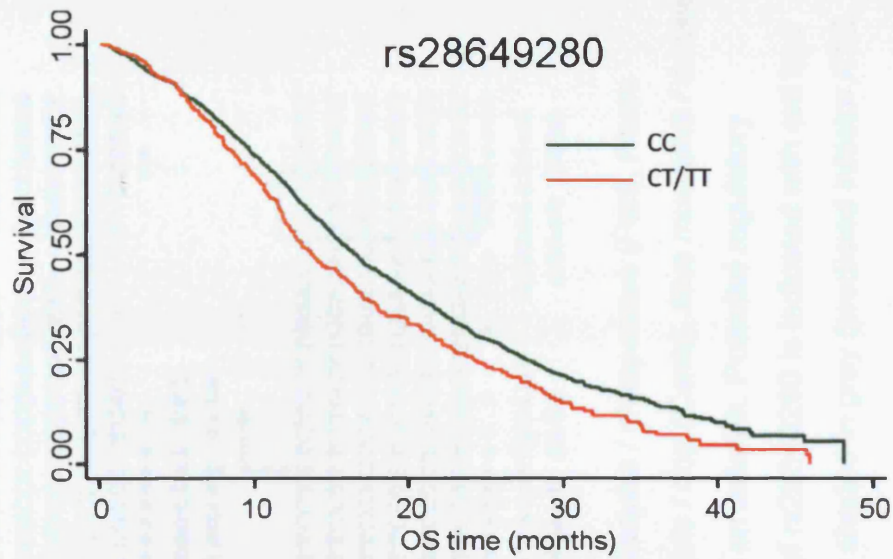
**Figure 4.6** - Kaplan-Meier curve depicting the effect of rs28649280 on OS. There was a difference in median OS of 2.9 months between the two groups.
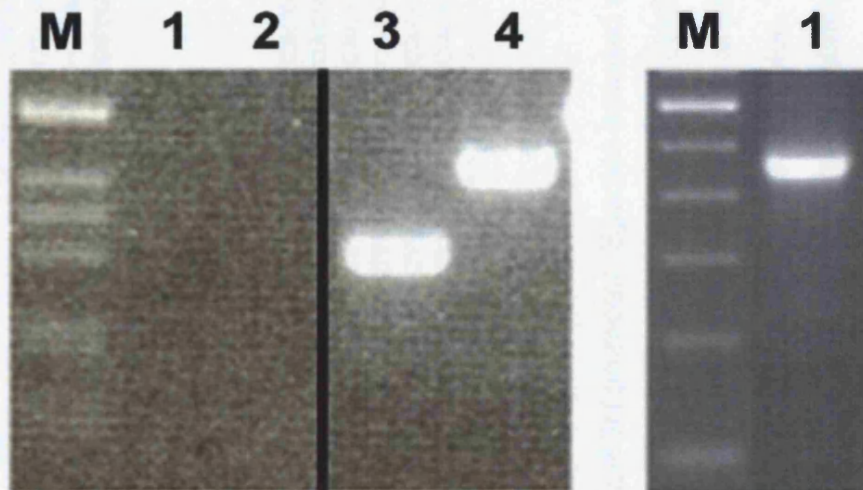


**Figure 4.7** - *Left*: 2% agarose gel of RT-PCR products using normal colonic cDNA as a template. Wells correspond to the reactions with the primers RT1F and RT3R (1), RT1F and RT4R (2), RT2F and RT3R (3), and RT2F and RT4R (4) as presented in Figure 4.4. M – 1kb marker. No products were generated with a forward primer in exon 1, indicating that this predicted exon was not transcribed within the colon. *Right*: 2% agarose gel of the 5'RACE product generated using RT4R. M – 500bp marker.

rs28649280

```
Pan      A-CTGGAAAACAGG---AGGGAGGCGGGGCCCAATTTTTGGCTTCTCCTTCCATCTCCGCCCCTCGAACGCC---GGCCTGCG
Homo     A-CTGGAAAACAGG---AGGGAGGCGGGGCCCAATTTTTGGCTTCTCCTTCCGTCTCCGCCCCTCGAACGCC---GGCCTGCG
Macaca   A-CTAGAGAAGAGG---AGGGAGGCGGGGCCCAATTTTTGGCTTCTCCTTCCATCTCCGCCCCCCGAACGCC---GGCCTGCG
Canis    AGCGGGACGAGCGG---AGGGAGGCGGGGCCCCGCGCCAGCCTCTCCTCCCG-CTCCGCCCCCCGGGCGCC---GGGCTGCT
Mus      AGGGAGAAACACCA-CCGAGGAGGCGGGGCCCTGCCCAGCC--TGCTTCCAGACACCGCCCCCCCCACCCCCAAAGCCTTAA
Rattus   AGGGGGAAACACCAACTGGGGAGGCGTGGCCCTACCCGGGCGGTGCTTCCAGACCCCGCCCCTCAC----------CCTTAC
         *      **           ******* *****        *        *   * *     * ******* *          **
```

                    Potential SP1                        Potential SP1
                    binding site                         binding site

```
Pan      AGTGACGTCACGCGAGACTTGTACTTCCGCTTCCGGTAACTCACGTTTCTCTTTCTTCCTGTCTGCTTGGAAAGATGGCGTC
Homo     AGTGACGTCACGCGAGACTTGTACTTCCGCTTCCGGTAACTCACGTTTCTCTTTCTTCCTGTCTGCTTGGAAAGATGGCGTC
Macaca   AGTGACGTCACGCGAGACTTGTACTTCCGCTTCCGGTAACTCACGTTTCTCTTTCTTCCTGTCAGCCCGGAAAGATGGCGTC
Canis    CGTGACGTCACGTGGAGTTCGTACTTCCGCTTCCGGCCACTCACGTTTCTCTTTTTTCCTGTCCGGCTGAAAAGATGGCGTC
Mus      CATAGCGTCACG-GAGCCTCGTAGTTCCGCTTCCGGCGCCTCACGCTTCTCTTTCTTCCTGTCTGGCTGGAAAGATGGCGTC
Rattus   GACAGCGTAA-G-GAACTTCGTAGTTCCGCTTCCGGCTCCTCACGCTTCTCTTTCTTCCTGTCTGGCTGGAAAGATGGCGTC
         ***  *  *       *  ***  ************  ******  ********  ******* *    *  *** ********
```

           Potential CREB          Potential Ets                    Transcription    Translation
           binding site            binding site                     start site       start site

**Figure 4.8** - Species conservation of the *EIF3H* promoter. Species included were *Pan troglodytes* / chimpanzee (*Pan*), *Homo sapiens* / human (*Homo*), *Macaca mulatta* / rhesus macaque (*Macaca*), *Canis lupus familiaris* / dog (*Canis*), *Mus musculus* / mouse (*Mus*), *Rattus norvegicus* / rat (*Rattus*). Bases identical in all six species are indicated with an asterisk. Potential regulatory elements are shaded; SP1 sites - green, CREB site – purple, Ets site - blue. The location of rs28649280 is indicated with red text and arrow. The transcription start site is coloured in scarlet and the translated sequence is shaded in grey (predicted initiator ATG boxed).

### 4.3.12 Luciferase expression analysis

A luciferase construct, containing a 136bp region upstream of the EIF3H transcription start site that included rs28649280, showed promoter activity as compared to empty vector ($P$=0.0021). This construct, containing the major C allele of rs28649280, was further compared against one containing the minor T allele (Figure 4.9). Construct containing the minor allele showed significantly increased expression levels as compared to that containing the major allele ($P$=0.0003).

### 4.3.13 *EIF3H* expression and protein analyses

We tested whether rs28649280 or rs16892766 affected *EIF3H* expression levels by examining RNA extracted from FFPE CRCs using two different *EIF3H* Taqman assays. Bland-Altman analysis of the data confirmed a strong correlation between the two assays. Although *EIF3H* was expressed in most tumours, we did not observe any differences in expression levels in CRCs from patients with the rs28649280-TT as compared to the -CC genotypes or from patients with the rs16892766-CC as compared to the -AA genotypes (Figures 4.10 & 4.11). EIF3H is thought to alter protein synthesis of several oncoproteins (Zhang *et al.* 2008); however, we did not find any differences in the levels of cyclin-D1, c-Myc, FGF-2 or in EIF3H itself, by immunohistochemical analysis of CRCs from patients with a range of rs28649280 and rs16892766 genotypes.
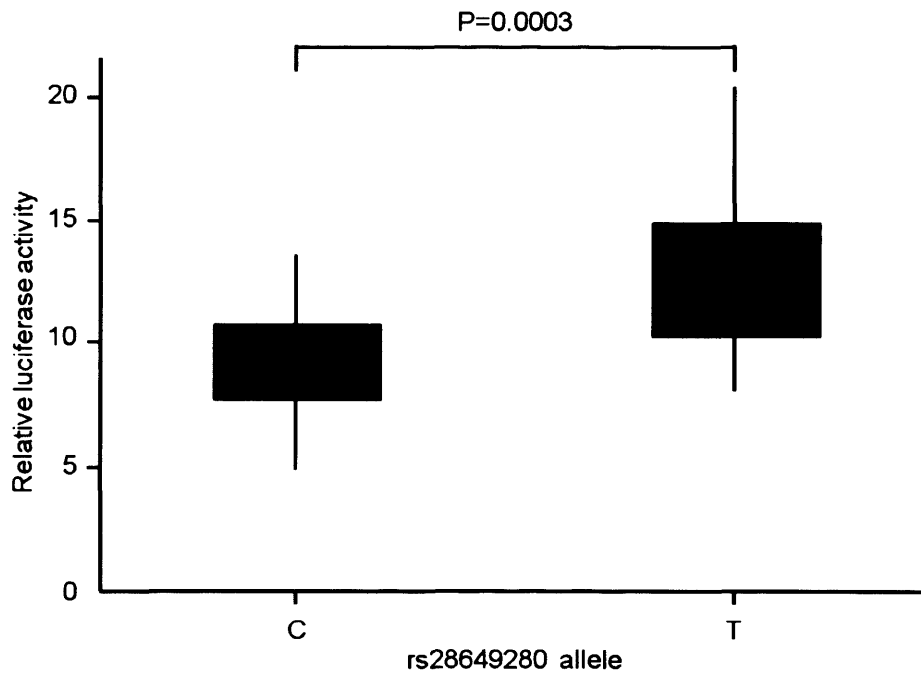
**Figure 4.9** – Box and whisker plot depicting the differences in expression levels between reporter construct containing the major C allele of rs28649280, and construct containing the minor T allele. There is a significantly increased relative luciferase activity in the latter (P=0.0003)
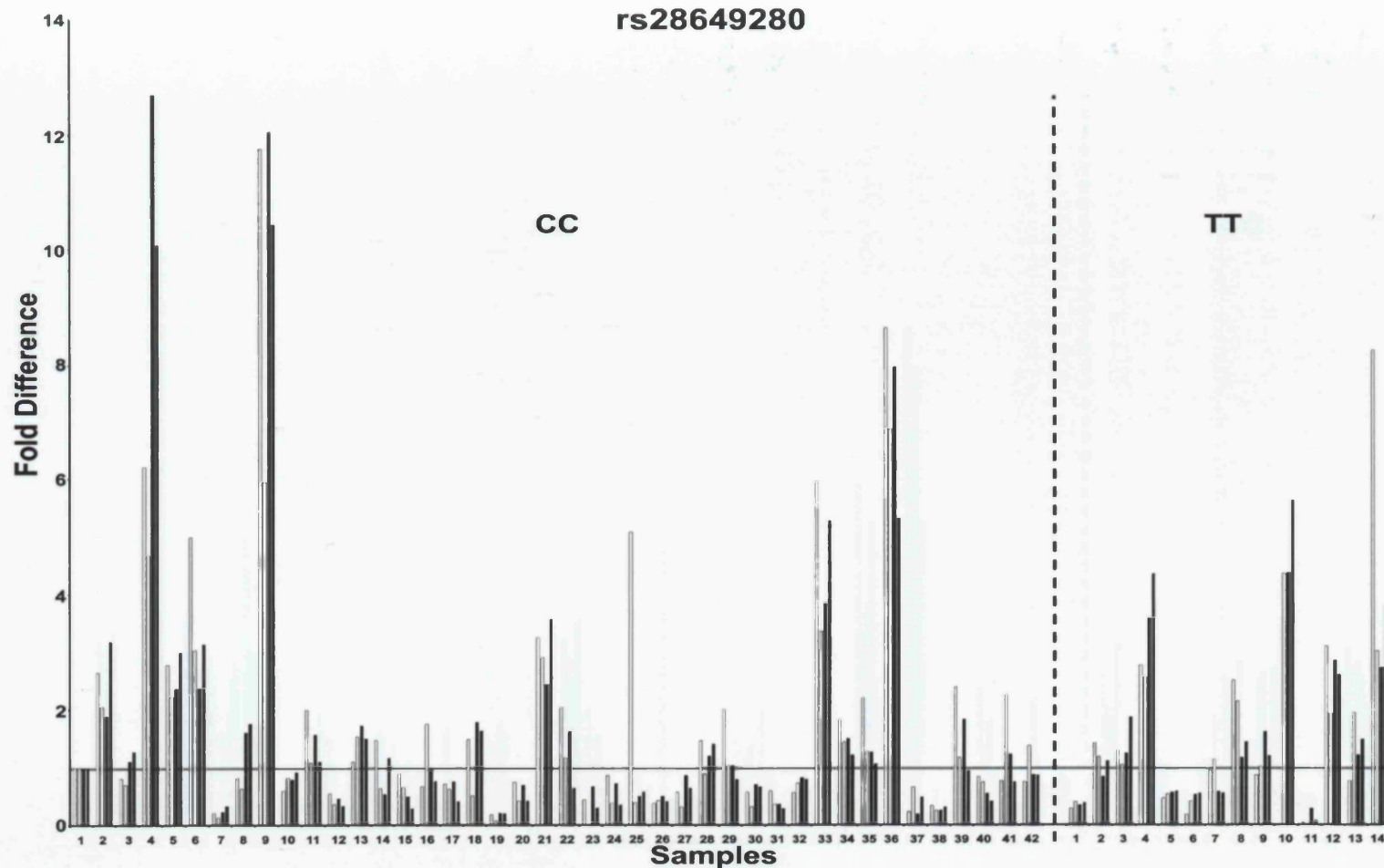
**Figure 4.10** - Bar chart of *EIF3H* expression levels in CRCs from patients with rs28649280-CC and -TT genotypes. Fold differences in expression levels were calculated relative to sample 1 (homozygous for the major allele). Each sample was analysed twice with two different assays (assay Hs01052033_m1, spanning exons 2 and 3 of *EIF3H*, is displayed in grey bars and assay Hs00186779_m1, spanning exons 6 and 8 of *EIF3H*, is displayed in black bars).
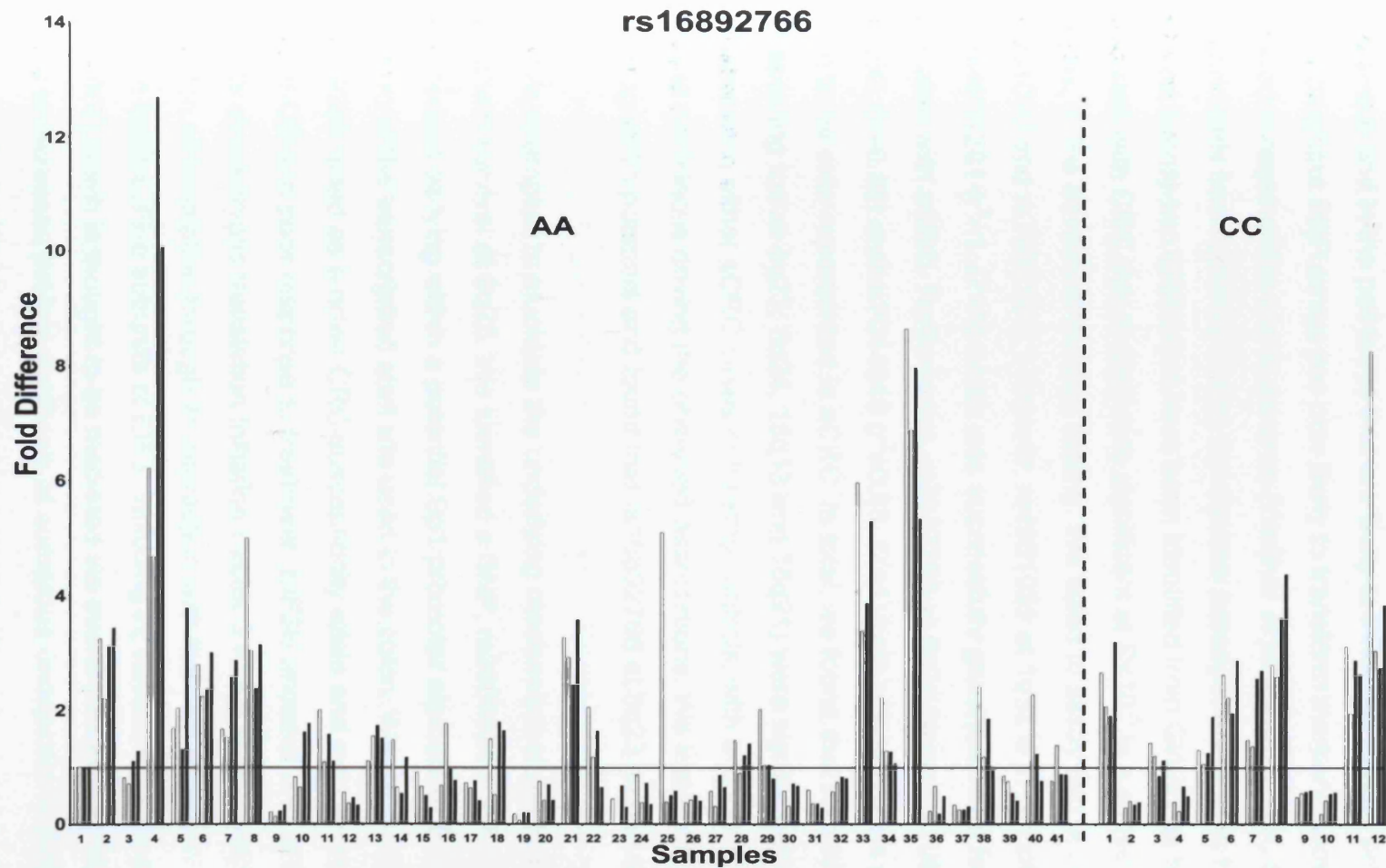
**Figure 4.11** - Bar chart of *EIF3H* expression levels in CRCs from patients with rs16892766-AA and -CC genotypes. Each sample was analysed twice with two different assays (see legend to Figure 4.10).

## 4.4 Discussion

To-date, the search for germline prognostic factors for CRC has focussed on the pharmacological pathways of the chemotherapeutic agents used in the treatment of this disease and in the pathways that are likely to influence tumour progression. High-throughput SNP arrays are now likely to transform marker discovery towards unbiased screens of the whole genome (Walther *et al.* 2009). So far, such arrays have primarily been used to search for disease susceptibility alleles. Here, we analysed twenty-two SNPs that have been identified from GWAS as being associated with CRC risk or that were significant at $P<10^{-5}$ in a subsequent meta-analysis, in the advanced disease setting. We failed to assay two of these SNPs (rs4951039 and rs7837328). However, rs4951039 at 1q32 is in complete association with rs4951291 ($r^2=1$, D'=1) which was successfully genotyped and found not to be associated with aCRC. Furthermore, rs7837328 at 8q24 tagged rs10808555 ($r^2=0.50$, D'=0.82) and rs7014346 ($r^2=0.83$, D'=1) both of which were genotyped and shown to be over-represented in aCRC. In total, we found that six SNPs (representing loci at 8q23, 8q24, 15q13 and 18q21) were significantly over-represented in either aCRC cases or healthy controls, with a single SNP of those tested at each locus driving the observed associations. We tested these SNPs for their prognostic potential and found that rs16892766 at 8q23 influenced OS.

We have attempted to elucidate the underlying mechanism of disease susceptibility and patient survival at 8q23. We identified a SNP, rs28649280, which was characterised as lying within a potential Sp1 promoter element of *EIF3H*, 78bp upstream of the transcription start site used in the colon. We showed that rs28649280 acted as a novel CRC-susceptibility allele and was associated with reduced OS and poor response to treatment. *EIF3H* encodes one of thirteen subunits of eukaryotic translation Initiation Factor 3 which is involved in the regulation of translation through its interaction with the 40S ribosome and other initiation factors. Five subunits of EIF3, including H, have oncogenic potential (Zhang *et al.* 2007) which is thought to be mediated via overexpression of these subunits leading to increased protein synthesis of numerous oncoproteins (Zhang *et al.* 2008).

EIF3H has been postulated to be responsible for polysome loading of mRNAs carrying upper ORFs (uORFs). These are found in the 5' leader sequences of genes

and inhibit translation initiation through a number of potential mechanisms (Sachs and Geballe, 2006). Critically, uORFs are found to cluster amongst mRNAs encoding regulatory proteins such as transcription factors and proto-oncogenes (Kozak, 1991). Current data suggests that EIF3H counteracts the inhibitory effects of uORFs, as well as preferentially boosting the translation of long mRNAs (Kim *et al.* 2007) in-turn defining a select group of mRNAs that undergo EIF3H dependant translation. Thus alterations in levels of EIF3H could conceivably lead to the altered translational state of numerous regulatory proteins. Indeed, such alterations have been observed for oncoproteins including cyclin D1, C-Myc, FGF2 and ornithine decarboxylase (Zhang *et al.* 2008).

We found that the minor allele of rs28649280 was associated with increased expression in a reporter gene assay system; however, we found no difference in the expression levels of *EIF3H* or in protein levels of FGF-2, C-myc and cyclin-D1 in CRCs from patients carrying the rs28649280 or rs16892766 risk alleles. Therefore, it is likely that the dysregulation of *EIF3H* expression and the downstream consequences that drive colorectal tumourigenesis, are masked by other cellular abnormalities that occur during progression to metastatic carcinoma. Further studies are therefore warranted to help unravel the mechanism through which *EIF3H* predisposes to CRC and affects patient outcome. Nonetheless, our findings raise the possibility that EIF3H may represent a suitable target for therapeutic intervention.

Though no studies have been conducted specifically targeting EIF3H, the related eukaryotic initiation factor 4E (EIF4E) has successfully been tested as a target in the pre-clinical setting. EIF4E binds the 5' cap of cellular mRNAs and delivers them to the eIF4F translation initiation complex. Like EIF3H, EIF4E enhances translation of a select set of mRNAs – including elements that are critical in regulating a number of cellular processes, which are in-turn up-regulated when *EIF4E* is over expressed (e.g. cyclin-D1 (Rosenwald *et al.* 1993) and C-myc (De Benedetti *et al.* 1994)). Over-expressed *EIF4E* ultimately leads to transformation of cells (Lazaris-Karatzas *et al.* 1990; Zimmer *et al.* 2000) and increased EIF4E expression has been linked with initiation and progression of specific human malignancies including breast (Li *et al.* 1997; Li *et al.* 1998), head and neck (Nathan *et al.* 1997), lung (Seki *et al.* 2002), and colorectal cancer (Rosenwald *et al.* 1999; Berkel *et al.* 2001). As shown by Graff and

colleagues (2007), *EIF4E* can be targeted with antisense oligonucleotides (ASO). ASO targeting of *EIF4E* mRNA led to repression of EIF4E regulated proteins including VEGF, cyclin D1, c-myc and ultimately led to suppression of tumour growth. Moreover, tests of the effects of ASOs on normal tissue showed that, despite a clear reduction in the levels of EIF4E, there was little to no adverse effects. Initial clinical trials of ASOs are currently ongoing. Agents that inhibit the interaction of EIF4E with one of its functional partners, EIF4G, are also currently being tested as novel therapies (Moerke *et al.* 2007; Tamburini *et al.* 2009). Other novel approaches include inhibitors of phosphorylation of EIF4E (Topisirovic *et al.* 2004) and the use of agents that interfere with the EIF4E-5' cap interaction (Kentsis *et al.* 2004; Assouline *et al.* 2009) have been employed with mixed success. It can be envisioned that similar approaches (Hsieh and Ruggero, 2010) could potentially be adopted in the therapeutic suppression of EIF3H in the treatment of CRC.

## 4.4.1 Future work

The functional mechanisms behind the 14 loci associated with CRC risk through GWAS have only just started to be understood. In the case of the association demonstrated at 8q23, further work is needed to confirm that EIF3H is indeed responsible for altering risk (and prognosis) of CRC. Recently, fine-mapping of the association signals at 8q23, 16q22 and 19q13, combined with data from expression quantitative trait loci (eQTL) databases suggest that the genes with the greatest biological plausibility for altering risk are not tagged by SNPs with the strongest association signal (Carvajal-Carmona *et al.* 2011). For example, the authors find that the SNPs with the greatest association signals at 8q23 alter expression of UTP23 and not EIF3H.

# Chapter Five - Multiple colorectal cancer susceptibility alleles affect patient survival and rs9929218 in E-cadherin predicts response to cetuximab

## 5.1 Introduction

As described in chapter 1, screens for inherited genetic factors that affect prognosis have suggested that pharmacological pathways may play an important role. However, the majority of prognostic biomarkers described to-date have not been confirmed in subsequent studies suggesting that many of the initial observations are false-positives due to the use of small, under-powered, sample sets and a lack of use of appropriate validation cohorts (Tejpar *et al.* 2010; Buyse *et al.* 2010). As discussed in chapter 4, one of the four validated CRC risk loci was shown to affect patient survival and response to treatment. Beyond this, no variants at the remaining loci have been shown to influence patient survival or to have a predictive effect on chemotherapy or drug response.

The treatment of aCRC is improving. Average survival has increased from ~6 months with best supportive care alone, through 10-12 months with 5FU-based regimens (Maughan *et al.* 2002) and up to 16-20 months with oxaliplatin and 5FU (Douillard *et al.* 2000, Saltz *et al.* 2000). In addition, cetuximab has been shown to improve OS in patients with CRC in whom other treatments have failed (Jonker *et al.* 2007). However, cetuximab efficacy is thought to be dependent upon an absence of somatic mutations in members of the EGFR signalling cascade such as *KRAS* and *BRAF* (De Roock *et al.* 2010). Despite these advances, around half of all patients still fail to respond to first line treatment (Allegra *et al.* 2009).

Here, we analysed SNPs from 12 CRC risk loci in a large collection of aCRC patients that had been recruited into the randomised controlled trials COIN and COIN-B. We tested for their prognostic effects on OS and response to, and side-effects from, chemotherapy, and for their predictive effects to cetuximab. We used independent training and validation cohorts to identify robust biomarkers and investigated the underlying mechanisms at 16q22.

## 5.2 Materials and Methods

### 5.2.1 Samples

We analysed 2,281 blood DNA samples from unrelated patients with aCRC from COIN (2,083 samples) and COIN-B (198 samples). For details of COIN patient clinicopathological features, and chemotherapeutic regimens applied, please refer to section 3.2.1, Figure 3.1 and Table 4.1. COIN-B patients were randomised 1:1 to receive intermittent oxaliplatin-based chemotherapy together with intermittent cetuximab (Arm D) and intermittent oxaliplatin-based chemotherapy together with continuous cetuximab (Arm E) (Figure 5.1). In all patients, treatment was identical for the first 12-weeks apart from the choice of oral capecitabine or infusional 5FU (patient/clinician choice) together with the randomisation of ± cetuximab. For the Arms with intermittent therapy, all treatment was stopped from 12-weeks (apart from continuous cetuximab in Arm E) if there was complete response (CR), partial response (PR) or stable disease (SD) and re-initiated upon disease progression (DP). There was no statistical difference in OS and PFS between any of the COIN Arms (Chapter 3, Maughan *et al.* 2011, Adams *et al.* 2011), and *KRAS, BRAF* and *NRAS* somatic mutation status did not affect response to cetuximab (Chapter 3; Maughan *et al.* 2011).

### 5.2.2 Genotyping

Genotyping of twelve previously identified CRC risk alleles (rs6691170 and rs6687758 at 1q41, rs10936599 at 3q26, rs4444235 and rs1957636 at 14q22, rs9929218 at 16q22, rs10411210 at 19q13, rs961253 at 20p12, rs10795668 at 10p14, rs3802842 at 11q23, rs4925386 at 20q13 and rs6983267 at 8q24) was performed by Illumina's Fast-Track Genotyping Services (San Diego, CA) using their high throughput BeadArray™ technology. Genotyping of rs4925386 at 20q13, rs16260 at 16q22, rs4813802 at 20p12 and, rs16969681 and rs11632715 at 15q13 was carried out by KBioscience (Herts, UK) using their KASPar technology. Genotyping of rs11169552 and rs7136702 at 12q13 was carried out by Geneservice (Nottingham, UK) using a TaqMan assay (Applied Biosystems).
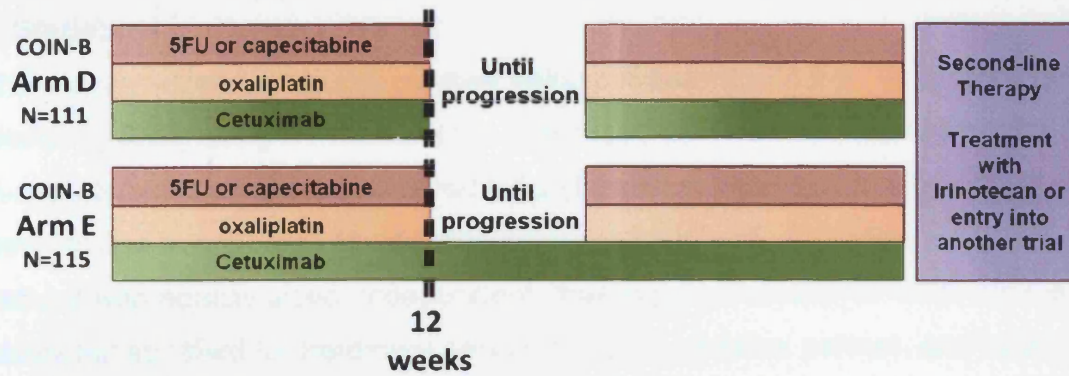
**Figure 5.1** - Treatment schedules for patients in COIN-B (COIN treatments are described in section 3.2.1 and Figure 3.1). COIN-B patients were randomised to receive intermittent oxaliplatin-based chemotherapy together with intermittent (Arm D) or continuous (Arm E) cetuximab. In all COIN and COIN-B patients, treatment was identical for the first 12-weeks apart from the choice of oral capecitabine or infusional 5FU (patient/clinician choice) together with the randomisation of ± cetuximab.

## 5.2.3 Statistical analyses

All SNPs were tested for their genotypes being consistent with the Hardy-Weinberg equilibrium (HWE) using a Pearson chi-square test. LD was examined using the open-source program Haploview version 4.2 (Barrett *et al.* 2005). To identify robust biomarkers and to minimise the identification of false-positives, all analyses were carried out with equally-sized, independent, 'training' and 'validation' sets, defined randomly but stratified for treatment arm and fluoropyrimidine partner, and balanced for all previously identified prognostic factors (WHO PS, number of metastatic sites, WBC count, alkaline phosphatase levels and *KRAS* and *BRAF* mutation status; Maughan *et al.* 2011). Significance levels were set at 5% ($P<0.05$) in both the training and validation phases, with effects in the same direction. Only SNPs significant in the training set were analysed in the validation set. SNPs passing both stages were tested by meta-analysis of the training and validation results using inverse-variance and fixed-effects. All analyses were adjusted for stratification factors and known prognostic factors. Two models were applied (i) the effect of at least one minor allele versus no minor alleles (Aa/aa vs. AA) and, (ii) the effect of two minor alleles versus one or no minor alleles (aa vs. Aa/AA).

For prognostic effects on OS, we combined then randomly split patients from all arms of COIN (OS data were not available from COIN-B) to form the training and validation sets. A Cox survival model was used. SNPs passing training and validation phases were re-analysed by treatment Arm to determine whether treatment influenced OS. All significant SNPs and known prognostic factors were entered into a closed-test procedure multiple fractional polynomial model with $P<0.05$.

All COIN and COIN-B patients had identical treatment regimes for the first 12 weeks apart from the addition of cetuximab (Figures 3.1 and 5.1). For prognostic effects on toxicity after 12 weeks of treatment, we combined then randomly split patients from Arms A and C of COIN (excluding those patients receiving cetuximab) to form the training and validation sets. For predictive effects on 12 week response to and side-effects from cetuximab, we combined then randomly split patients from Arm B of COIN with all of those from COIN-B (standard chemotherapy plus cetuximab) and

compared these groups to patients similarly combined then split from Arms A and C of COIN (standard chemotherapy without cetuximab).

Twelve-week response was defined as PR or CR versus SD, PD or a blank field in the clinical return form (CRF), recorded between 56 and 112 days (8 and 16 weeks) from randomisation. If no CRFs were available between those time limits, then patients were excluded from the analysis. Twelve-week toxicity was defined as the maximum toxicity recorded in any CRF returned within 112 days (16 weeks) of randomisation. Definitions of all other toxicities have been previously described in section 4.2.8. All prognostic and predictive analyses were carried out by David Fisher.

## 5.3 Results

### 5.3.1 Genotyping

We genotyped seventeen CRC-risk alleles identified from GWAS (Tomlinson *et al.* 2007, Broderick *et al.* 2007, Jaeger *et al.* 2008, Tenesa *et al.* 2008, Tomlinson *et al.* 2008, Houlston *et al.* 2008, Houlston *et al.* 2010, Carvajal-Carmona *in submission*), representing 12 loci (a single SNP at seven loci and two SNPs at five loci; Houlston *et al.* 2008, Houlston *et al.* 2010, Carvajal-Carmona *in submission*), in 2,281 unrelated patients from COIN and COIN-B using three different platforms. Twelve SNPs were genotyped using the Illumina GoldenGate platform of which one (rs4925386) failed genotyping. For the remaining 11 SNPs, genotyping concordance rates for duplicate samples was 100% (1,210/1,210), GenTrain scores ranged from 0.6814 to 0.9500 and the overall genotyping success rate was 99.34% (23,981/24,141) genotypes were called successfully). Four SNPs, rs4925386, rs4813802, rs16969681 and rs11632715 were genotyped using KASPar technology with a genotype success rate of 98.88% (8,928/9,029 genotypes called successfully) and concordance rate for duplicate samples of 100% (376/376). Two SNPs, rs11169552 and rs7136702 were genotyped using Taqman assays with a genotype success rate of 95.52% (4176/4372 genotypes called successfully) and concordance rate for duplicate samples of 100% (188/188). All SNPs, apart from rs7136702 (P=0.027), had genotype distributions consistent with the HWE and we found no imbalances in the frequencies of the SNPs between the treatment arms or according

to the somatic *KRAS*, *BRAF* and *NRAS* mutation status of the CRCs (42.27%, 9.01% and 3.56% of the CRCs were *KRAS*, *BRAF* and *NRAS* mutant, respectively; Maughan *et al.* 2011).

### 5.3.2 Prognostic effects on patient survival

We have previously reported on the prognostic and predictive effects of rs16892766, rs4939827 and rs4779584, as discussed in chapter 4. We did not observe any effect of rs6983267, rs11169552, rs6691170, rs6687758, rs10936599, rs4444235, rs961253, rs3802842, rs1957636, rs16969681 or rs11632715 on OS within the training phase. Although rs7136702, rs4925386 and rs4813802 were significant in the training phase, all failed validation (Table 5.1). In contrast, rs9929218, rs10411210 and rs10795668 significantly influenced OS within both the training and validation phases under a univariate model (Table 5.1). After meta-analysis of the combined data, we found that patients homozygous for the minor allele (AA) of rs9929218 (equating to ~8% of patients) showed a significant decline in OS as compared to patients with the AG or GG genotypes (HR 1.46, 95% CI 1.23–1.74, P=0.002 after correction for heterogeneity), with a median decrease in life expectancy of 4.3 months (Table 5.1, Figure 5.2). Similarly, for rs10411210, patients carrying the TC or TT genotype had worse OS as compared to those with the CC genotype (HR 1.23, 95% CI 1.08–1.39, P=0.001); median decrease in life expectancy of 2.5 months. For rs10795668, patients carrying the AA genotype had an improved OS as compared to those with the AG or GG genotypes (HR 0.71, 0.59–0.86, P=0.001); median increase in life expectancy of 6.2 months.

To determine whether these findings were influenced by the type and duration of treatment, we evaluated these variants according to trial Arm. For rs9929218 and rs10411210 the effects were significant across Arms A and B (Table 5.2); however for patients treated with intermittent oxaliplatin and fluoropyrimidine chemotherapy (Arm C), the HRs did not reach statistical significance, suggesting that these alleles may influence survival based upon an interaction with oxaliplatin and fluoropyrimidine treatment. In contrast, the effect of rs10795668 was most significant in patients from Arm C (HR 0.64, 95% CI 0.47–0.88, P=0.005), suggesting that this prognostic factor was likely to be independent of treatment (Table 5.2).

We have previously shown that the WHO PS, number of metastatic sites, WBC count, alkaline phosphatase levels and *KRAS* and *BRAF* mutation status are independent prognostic factors affecting OS in patients from COIN (Maughan *et al.* 2011). We therefore applied a multivariate model including rs9929218, rs10411210 and rs10795668 and the known prognostic factors, and showed that all of these SNPs independently influenced OS.

It is possible that the drop-out of associations from the training to validation phases might be due to the phenomenon known as the winner's curse. This states that in genetic association studies there is likely to be an over-estimation of true genetic effect sizes (Xiao and Boehnke, 2009). This occurs as a result of ascertainment bias; that is, a bias resulting from the selection of only those associations that are significant in the initial data set for further analysis. As such, replication cohorts are likely to be underpowered to detect the same association.

| | Genotype(s) | n | Training phase HR (95% CI) | P-value | n | Validation phase HR (95% CI) | P-value* | Combined HR (95% CI) | P-value | $I^2$ | RE P-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs6983267 | TT/TG | 689 | 1.02 (0.88–1.19) | 0.755 | - | - | - | - | - | - | - |
| | TT | 220 | 1.15 (0.97–1.36) | 0.104 | - | - | - | - | - | - | - |
| rs11169552 | CT/TT | 479 | 1.01 (0.88–1.16) | 0.889 | - | - | - | - | - | - | - |
| | TT | 76 | 1.11 (0.85–1.45) | 0.431 | - | - | - | - | - | - | - |
| rs7136702 | CT/TT | 577 | 1.05 (0.91–1.22) | 0.489 | - | - | - | - | - | - | - |
| | TT | 135 | 1.30 (1.06–1.50) | 0.012 | 154 | 1.04 (0.85–1.26) | 0.726 | - | - | - | - |
| rs6691170 | TG/TT | 682 | 0.92 (0.80–1.07) | 0.293 | - | - | - | - | - | - | - |
| | TT | 163 | 0.93 (0.77–1.13) | 0.485 | - | - | - | - | - | - | - |
| rs6687758 | AG/GG | 400 | 0.90 (0.78–1.04) | 0.171 | - | - | - | - | - | - | - |
| | GG | 55 | 1.04 (0.78–1.38) | 0.801 | - | - | - | - | - | - | - |
| rs10936599 | TC/TT | 431 | 1.08 (0.94–1.24) | 0.300 | - | - | - | - | - | - | - |
| | TT | 57 | 1.33 (0.99–1.79) | 0.059 | - | - | - | - | - | - | - |
| rs4925386 | TC/TT | 534 | 0.86 (0.75–0.99) | 0.034 | 554 | 0.98 (0.85–1.13) | 0.732 | - | - | - | - |
| | TT | 91 | 1.00 (0.77–1.29) | 0.984 | - | - | - | - | - | - | - |
| rs4444235 | TC/CC | 754 | 1.02 (0.87–1.19) | 0.792 | - | - | - | - | - | - | - |
| | CC | 250 | 0.97 (0.82–1.15) | 0.737 | - | - | - | - | - | - | - |
| rs9929218 | AG/AA | 520 | 1.02 (0.89–1.18) | 0.736 | - | - | - | - | - | - | - |
| | AA | 91 | 1.30 (1.03–1.65) | 0.028 | 73 | 1.68 (1.29–2.17) | <0.001 | 1.46 (1.23–1.74) | <0.001 | 49% | 0.002 |
| rs10411210 | TC/TT | 191 | 1.23 (1.02–1.46) | 0.026 | 193 | 1.23 (1.03–1.47) | 0.022 | 1.23 (1.08–1.39) | 0.001 | 0% | 0.001 |
| | TT | 12 | 1.02 (0.53–1.97) | 0.959 | - | - | - | - | - | - | - |
| rs961253 | AC/AA | 628 | 1.07 (0.93–1.24) | 0.338 | - | - | - | - | - | - | - |
| | AA | 143 | 0.99 (0.81–1.21) | 0.895 | - | - | - | - | - | - | - |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs10795668 | AG/AA | 509 | 0.89 (0.77–1.02) | 0.094 | - | - | - | - | - | - | - |
| | AA | 89 | 0.70 (0.54–0.92) | 0.011 | 96 | 0.72 (0.55–0.95) | 0.018 | 0.71 (0.59–0.86) | 0.001 | 0% | 0.001 |
| rs3802842 | AC/CC | 564 | 1.00 (0.87–1.15) | 0.973 | - | - | - | - | - | - | - |
| | CC | 108 | 1.10 (0.87–1.37) | 0.431 | - | - | - | - | - | - | - |
| rs4813802 | GT/GG | 718 | 1.02 (0.87–1.18) | 0.843 | - | - | - | - | - | - | - |
| | GG | 194 | 0.80 (0.67–0.96) | 0.017 | 190 | 1.14 (0.95–1.37) | 0.148 | - | - | - | - |
| rs1957636 | GA/GG | 638 | 0.93 (0.81–1.08) | 0.332 | - | - | - | - | - | - | - |
| | GG | 143 | 1.15 (0.93–1.41) | 0.189 | - | - | - | - | - | - | - |
| rs16969681 | TC/TT | 208 | 1.07 (0.90–1.27) | 0.447 | - | - | - | - | - | - | - |
| | TT | 16 | 1.46 (0.89–2.40) | 0.134 | - | - | - | - | - | - | - |
| rs11632715 | GA/GG | 779 | 0.86 (0.74–1.01) | 0.072 | - | - | - | - | - | - | - |
| | GG | 264 | 1.14 (0.98–1.34) | 0.099 | - | - | - | - | - | - | - |

**Table 5.1** - Univariate prognostic analyses of OS. Analyses use Cox proportional-hazards models with the outcome of OS. Hazard ratios should be interpreted as the hazard in the genotype group indicated versus all other patients with a result for that SNP. *Correction for multiple testing was applied at this stage, though data is not shown. rs9929218 was the only SNP to remain significant (P<0.006) after this correction.
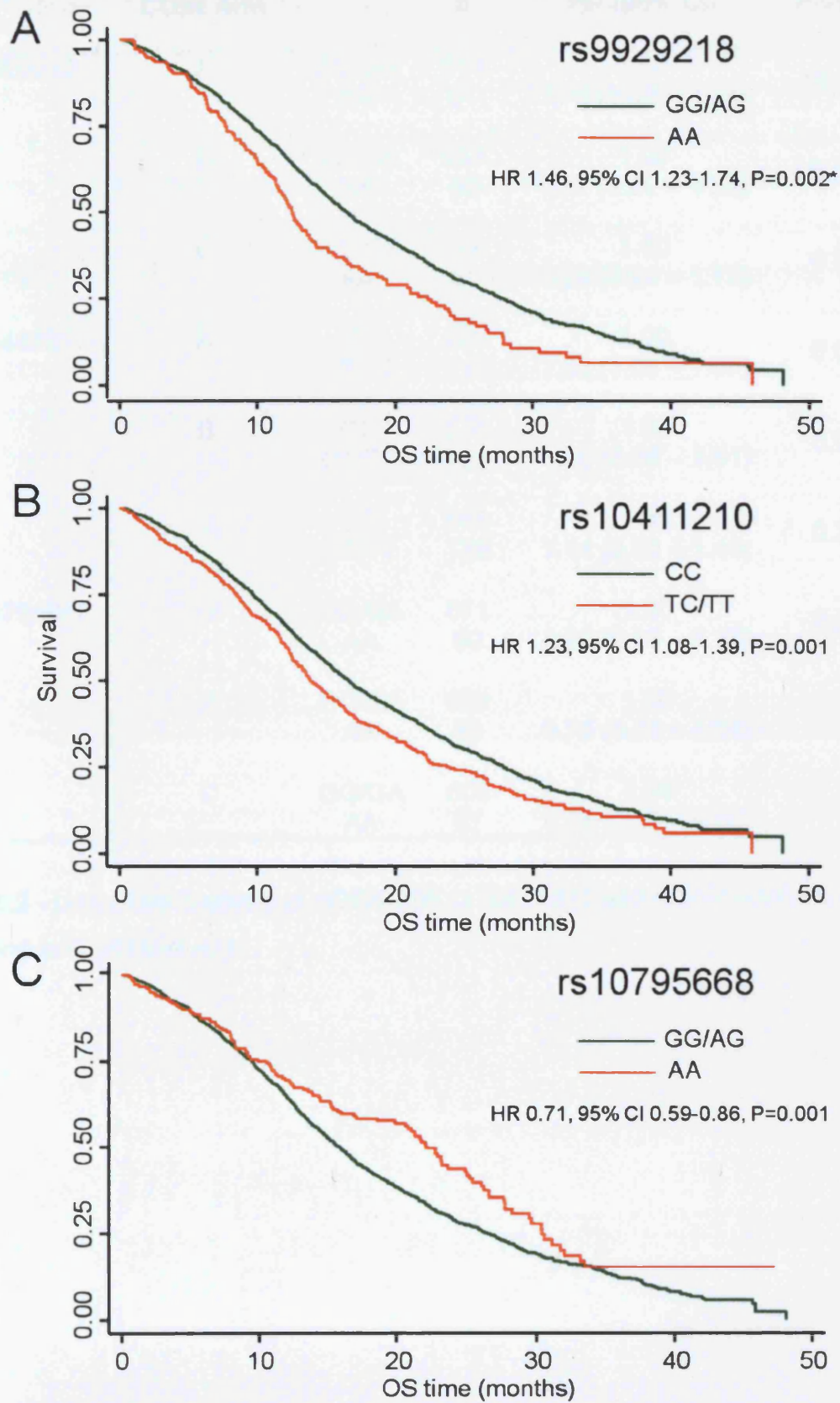
**Figure 5.2 -** Kaplan-Meier curves for the effect of (A) rs9929218, (B) rs10411210 and (C) rs10795668 on OS. *after correction for heterogeneity.

| | COIN Arm | | n | HR (95% CI) | P-value |
|---|---|---|---|---|---|
| rs9929218 | A | GG/GA | 611 | 1.00 | ≤0.001 |
| | | AA | 61 | 1.76 (1.32 – 2.34) | |
| | B | GG/GA | 657 | 1.00 | 0.048 |
| | | AA | 52 | 1.37 (1.00 – 1.88) | |
| | C | GG/GA | 646 | 1.00 | 0.116 |
| | | AA | 51 | 1.29 (0.94 – 1.77) | |
| rs10411210 | A | CC | 558 | 1.00 | 0.013 |
| | | CT/TT | 113 | 1.33 (1.06 – 1.68) | |
| | B | CC | 570 | 1.00 | 0.071 |
| | | CT/TT | 132 | 1.22 (0.98 – 1.51) | |
| | C | CC | 558 | 1.00 | 0.241 |
| | | CT/TT | 139 | 1.14 (0.92 – 1.40) | |
| rs10795668 | A | GG/GA | 571 | 1.00 | 0.248 |
| | | AA | 60 | 0.82 (0.59 – 1.15) | |
| | B | GG/GA | 624 | 1.00 | 0.082 |
| | | AA | 58 | 0.73 (0.52 – 1.04) | |
| | C | GG/GA | 605 | 1.00 | 0.005 |
| | | AA | 67 | 0.64 (0.47 – 0.88) | |

**Table 5.2** - Univariate analysis of rs9929218, rs10411210 and rs10795668 on OS according to COIN trial Arm

### 5.3.3 Prognostic effects on response to, or side effects from, chemotherapy

We tested the CRC-susceptibility SNPs for their effects on response to, or side effects from, 12 weeks of fluoropyrimidine and oxaliplatin-based chemotherapy (Table 5.3). We found that patients homozygous for the rs9929218 minor allele had a poor prognosis for response to treatment in both the training and validation phases (combined OR 0.48, 95% CI 0.31–0.75, P=0.001), as compared to patients that were heterozygous or homozygous wild-type (Table 5.4). In addition, we found that this subgroup of patients had an increased risk of skin toxicity, specifically HFS within the training phase (OR 6.04, 95% CI 1.65-22.2, P=0.007), that strongly trended towards significance in the validation phase (OR 3.55, 95% CI 0.92-13.7, P=0.065) and that was highly significant in the combined dataset (OR 4.68, 95% CI 1.84–11.9, P=0.001) (Table 5.4).

| | | Response | | GI toxicity | | Haematological toxicity | | Skin toxicity | |
|---|---|---|---|---|---|---|---|---|---|
| | | TP | VP* | TP | VP* | TP | VP* | TP | VP* |
| rs6983267 | TG/TT | 1.07 (0.76–1.52) P=0.681 | - | 1.21 (0.76–1.93) P=0.423 | - | 1.43 (0.8–2.55) P=0.229 | - | 0.42 (0.12–1.39) P=0.154 | - |
| | TT | 0.73 (0.49–1.09) P=0.122 | - | 0.81 (0.47–1.39) P=0.446 | - | 1.54 (0.85–2.81) P=0.154 | - | 2.12 (0.61–7.39) P=0.239 | - |
| rs11169552 | CT/TT | 1.11 (0.79–1.55) P=0.555 | - | 1.17 (0.76–1.81) P=0.477 | - | 1.3 (0.76–2.23) P=0.344 | - | 1.46 (0.44–4.84) P=0.541 | - |
| | TT | 0.97 (0.50–1.87) P=0.923 | - | 1.01 (0.46–2.21) P=0.974 | - | 0.96 (0.35–2.68) P=0.944 | - | 1.34 (0.17–10.8) P=0.786 | - |
| rs7136702 | CT/TT | 0.89 (0.63–1.25) P=0.489 | - | 1.11 (0.71–1.72) P=0.652 | - | 0.71 (0.41–1.23) P=0.224 | - | 0.85 (0.25–2.82) P=0.785 | - |
| | TT | 0.98 (0.61–1.59) P=0.950 | - | 1.21 (0.64–2.29) P=0.556 | - | 0.97 (0.44–2.14) P=0.940 | - | 2.47 (0.64–9.59) P=0.191 | - |
| rs6691170 | TG/TT | 0.86 (0.61–1.22) P=0.388 | - | 1.10 (0.70–1.73) P=0.693 | - | 1.09 (0.63–1.92) P=0.752 | - | 2.42 (0.52–11.3) P=0.262 | - |
| | TT | 0.85 (0.51–1.39) P=0.512 | - | 1.36 (0.71–2.58) P=0.352 | - | 1.13 (0.49–2.59) P=0.780 | - | 5.38 (1.49–19.5) P=0.010 | 0.95 (0.21–4.36) P=0.952 |
| rs6687758 | AG/GG | 0.92 (0.66–1.29) P=0.638 | - | 0.65 (0.41–1.03) P=0.066 | - | 0.87 (0.50–1.52) P=0.625 | - | 2.87 (0.83–9.94) P=0.096 | - |
| | GG | 0.62 (0.29–1.32) P=0.214 | - | 1.47 (0.55–3.95) P=0.446 | - | 1.66 (0.51–5.39) P=0.401 | - | 2.76 (0.33–23.0) P=0.348 | - |
| rs10936599 | TC/TT | 1.08 (0.77–1.51) P=0.647 | - | 0.94 (0.61–1.45) P=0.780 | - | 0.81 (0.47–1.39) P=0.434 | - | 0.50 (0.13–1.93) P=0.317 | - |
| | TT | 1.40 (0.64–3.08) P=0.397 | - | 0.16 (0.02–1.19) P=0.074 | - | 0.38 (0.08–1.72) P=0.210 | - | No pts. | - |
| rs4925386 | TC/TT | 1.11 (0.80–1.55) P=0.524 | - | 1.02 (0.66–1.58) P=0.933 | - | 1.14 (0.66–1.95) P=0.644 | - | 0.94 (0.28–3.15) P=0.923 | - |
| | TT | 1.29 (0.75–2.23) P=0.353 | - | 1.04 (0.52–2.06) P=0.916 | - | 0.79 (0.33–1.87) P=0.59 | - | 1.73 (0.36–8.27) P=0.494 | - |

141

| | | Response | | GI toxicity | | Haematological toxicity | | Skin toxicity | |
|---|---|---|---|---|---|---|---|---|---|
| | | TP | VP* | TP | VP* | TP | VP* | TP | VP* |
| rs4444235 | TC/CC | 1.01 (0.70–1.46) P=0.944 | - | 1.18 (0.71–1.94) P=0.520 | - | 0.92 (0.51–1.66) P=0.788 | - | 1.00 (0.26–3.84) P=0.999 | - |
| | CC | 0.71 (0.49–1.05) P=0.084 | - | 0.93 (0.57–1.51) P=0.762 | - | 0.70 (0.36–1.36) P=0.295 | - | 1.79 (0.52–6.23) P=0.359 | - |
| rs9929218 | AG/AA | 1.10 (0.79–1.53) P=0.564 | - | 1.10 (0.72–1.69) P=0.661 | - | 1.20 (0.71–2.05) P=0.495 | - | 1.79 (0.52-6.21) P=0.356 | - |
| | AA | 0.49 (0.27–0.89) P=0.019 | 0.47 (0.24–0.91) P=0.026 | 1.02 (0.50–2.08) P=0.962 | - | 0.92 (0.36–2.32) P=0.852 | - | 7.59 (2.23–25.9) P=0.001 | 3.35 (0.88–12.7) P=0.076[t] |
| rs10411210 | TC/TT | 1.05 (0.66–1.68) P=0.840 | - | 0.97 (0.51–1.85) P=0.931 | - | 1.93 (1.00–3.73) P=0.050 | 1.35 (0.74–2.45) P=0.330 | 1.44 (0.30–6.90) P=0.645 | - |
| | TT | 1.09 (0.18–6.57) P=0.926 | - | 11.2 (0.94–132) P=0.056 | - | No pts. | - | 15.1 (1.46–156) P=0.023 | No pts. |
| rs961253 | AC/AA | 0.88 (0.63–1.23) P=0.447 | - | 1.01 (0.65–1.56) P=0.972 | - | 0.57 (0.33–0.97) P=0.039 | 1.20 (0.70–2.07) P=0.506 | 3.13 (0.67–14.7) P=0.148 | - |
| | AA | 2.15 (1.25–3.70) P=0.006 | 1.22 (0.76–1.96) P=0.404 | 0.53 (0.24–1.16) P=0.110 | - | 0.13 (0.03–0.57) P=0.007 | 1.26 (0.62–2.53) P=0.522 | 1.77 (0.37–8.41) P=0.475 | - |
| rs10795668 | AG/AA | 0.96 (0.68–1.34) P=0.791 | - | 1.33 (0.84–2.08) P=0.222 | - | 1.57 (0.89–2.75) P=0.117 | - | 2.25 (0.59–8.60) P=0.237 | - |
| | AA | 1.33 (0.75–2.34) P=0.332 | - | 1.18 (0.60–2.32) P=0.629 | - | 0.64 (0.24–1.75) P=0.385 | - | No pts. | - |
| rs3802842 | AC/CC | 0.98 (0.71–1.36) P=0.911 | - | 1.01 (0.66–1.55) P=0.969 | - | 1.03 (0.60–1.76) P=0.915 | - | 0.74 (0.22–2.47) P=0.628 | - |
| | CC | 1.24 (0.67–2.29) P=0.486 | - | 0.81 (0.34–1.92) P=0.638 | - | 0.78 (0.26–2.35) P=0.658 | - | No pts. | - |
| rs4813802 | GT/GG | 0.72 (0.50–1.04) P=0.081 | - | 1.22 (0.75–1.98) P=0.423 | - | 0.80 (0.46–1.4) P=0.436 | - | 4.31 (0.55–34.0) P=0.166 | - |
| | GG | 1.05 (0.68–1.63) P=0.830 | - | 0.81 (0.45–1.46) P=0.492 | - | 0.48 (0.2–1.19) P=0.114 | - | 4.05 (1.21–13.6) P=0.024 | 0.69 (0.15–3.13) P=0.630 |

| | | Response | | GI toxicity | | Haematological toxicity | | Skin toxicity | |
|---|---|---|---|---|---|---|---|---|---|
| | | TP | VP* | TP | VP* | TP | VP* | TP | VP* |
| rs1957636 | GA/GG | 0.99 (0.70–1.39) P=0.946 | - | 0.95 (0.61–1.46) P=0.804 | - | 0.77 (0.45–1.33) P=0.345 | - | 0.79 (0.24–2.64) P=0.707 | - |
| | GG | 0.82 (0.52–1.28) P=0.372 | - | 1.08 (0.61–1.91) P=0.792 | - | 0.72 (0.32–1.63) P=0.434 | - | 1.24 (0.26–5.86) P=0.785 | - |
| rs16969681 | TC/TT | 0.89 (0.59–1.33) P=0.556 | - | 0.75 (0.43–1.29) P=0.296 | - | 0.70 (0.34–1.45) P=0.335 | - | 0.75 (0.16–3.54) P=0.717 | - |
| | TT | 1.28 (0.37–4.42) P=0.696 | - | 3.08 (0.65–14.6) P=0.156 | - | No pts. | - | No pts. | - |
| rs11632715 | GA/GG | 1.09 (0.75–1.58) P=0.666 | - | 1.05 (0.64–1.72) P=0.854 | - | 1.62 (0.85–3.08) P=0.145 | - | 0.38 (0.11–1.28) P=0.120 | - |
| | GG | 0.95 (0.65–1.39) P=0.785 | - | 0.89 (0.53–1.49) P=0.663 | - | 1.15 (0.61–2.17) P=0.665 | - | 0.66 (0.14–3.12) P=0.603 | - |

**Table 5.3** - Prognostic effects of CRC-risk SNPs on response to and side effects from 12 weeks of oxaliplatin and fluoropyrimidine based chemotherapy. Numbers represent HRs with 95% CI in parenthesese. Hazard ratios should be interpreted as the hazard in the genotype group indicated versus all other patients with a result for that SNP. TP - Training phase, VP – Validation phase. Significant results in both training and validation phases are shaded ([t]N.B. Patients carrying the AA genotype for rs9929218 were marginally over the 5% significance level for worse skin toxicity within the validation phase, but were highly significant in the subsequent meta-analyses [P<0.001]). No pts. – no patients in that phase with that genotype. *Though data is not shown, correction for multiple testing was applied after the VP. All observations were non-significant after this correction, though rs9929218 showed border-line significance (P=0.052) for prognostic effect on response.

| | | Training phase | | | Validation phase | | | Combined | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | n | OR (95%CI) | P-value | n | OR (95%CI) | P-value | OR (95%CI) | P-value | $i^2$ | RE P-value |
| 12-week response | AG/AA | 279 | 1.10 (0.79 – 1.53) | 0.564 | - | - | - | - | - | - | - |
| | AA | 50 | 0.49 (0.27 – 0.89) | 0.019 | 40 | 0.47 (0.24 – 0.91) | 0.026 | 0.48 (0.31 – 0.75) | 0.001 | 0% | 0.001 |
| 12-week GI toxicity | AG/AA | 257 | 1.10 (0.72 – 1.69) | 0.661 | - | - | - | - | - | - | - |
| | AA | 50 | 1.02 (0.50 – 2.08) | 0.962 | - | - | - | - | - | - | - |
| 12-week haemato-logical toxicity | AG/AA | 274 | 1.20 (0.71 – 2.05) | 0.495 | - | - | - | - | - | - | - |
| | AA | 55 | 0.92 (0.36 – 2.32) | 0.852 | - | - | - | - | - | - | - |
| 12-week skin toxicity | AG/AA | 306 | 1.79 (0.52 – 6.21) | 0.356 | - | - | - | - | - | - | - |
| | AA | 62 | 7.59 (2.23 – 25.9) | 0.001 | 44 | 3.35 (0.88 – 12.7) | 0.076 | 5.22 (2.12 – 12.9) | <0.001 | 0% | <0.001 |
| 12 week HFS | AG/AA | 311 | 1.55 (0.43 – 5.57) | 0.500 | - | - | - | - | - | - | - |
| | AA | 63 | 6.04 (1.65 – 22.2) | 0.007 | 44 | 3.55 (0.92 – 13.7) | 0.065 | 4.68 (1.84 – 11.9) | 0.001 | 0% | 0.001 |

**Table 5.4** - Prognostic effect of rs9929218 in E-cadherin on response to and side effects from 12 weeks of oxaliplatin and fluoropyrimidine based chemotherapy. Odds ratios should be interpreted as the odds in the genotype group indicated versus all other patients with a result for that SNP.

144

We investigated whether the somatic *KRAS*, *BRAF* and *NRAS* status of the CRCs influenced the prognostic effect of rs9929218 on response to treatment. In patients with *KRAS*, *BRAF* or *NRAS* mutant tumours, those who were homozygous for the rs9929218 minor allele in the germline showed worse response than those heterozygous or homozygous for the wild type allele (OR 0.25, 95% CI 0.11-0.55, *P*=0.001). In contrast, amongst patients with *KRAS*, *BRAF* and *NRAS* wild type tumours, those who were homozygous for the rs9929218 minor allele did not have worse response than those heterozygous or homozygous for the wild type allele (OR 0.85, 95% CI 0.40-1.83, *P*=0.69) The difference in response for patients homozygous for the rs9929218 minor allele with respect to somatic mutation status was significant (*P* for interaction 0.037).

### 5.3.4 Predictive effects on response to cetuximab

We tested whether any of the SNPs might also act as predictive factors for response to cetuximab (Table 5.5). rs9929218 was found to predict response within both the training and validation phases. After meta-analysis of the combined data, patients homozygous for the rs9929218 minor allele showed a significantly improved response (OR 1.69, 95% CI 0.93-3.07 with cetuximab, OR 0.48, 95% CI 0.31-0.75 without for cetuximab; P for interaction = 0.001) (Table 5.6). Although an equivalent effect on OS was not observed (HR 1.38, 95% CI 1.01-1.89 with cetuximab treatment and HR 1.53, 95% CI 1.24-1.89 without cetuximab treatment; P for interaction 0.47), there was considerable heterogeneity in treatments after 12 weeks.

Patients homozygous for the rs9929218 minor allele also had reduced skin toxicity after 12 weeks of cetuximab treatment in the training and validation phases (combined OR 0.72, 95% CI 0.32-1.61 with cetuximab, OR 5.22, 95% CI 2.12-12.9 without cetuximab; P for interaction <0.001) (Table 3), specifically, reduced HFS (combined OR 0.84, 95% CI 0.25-2.83 with cetuximab, OR 4.68, 95% CI 1.84-11.9 without cetuximab; P for interaction 0.030) (Table 3).

| | | Response | | GI toxicity | | Haematological toxicity | | Skin toxicity | |
|---|---|---|---|---|---|---|---|---|---|
| | | TP | VP* | TP | VP* | TP | VP* | TP | VP* |
| rs6983267 | TG/TT | 1.06 (0.59–1.91) P=0.849 | - | 0.78 (0.38–1.62) P=0.508 | - | 1.2 (0.46–3.15) P=0.706 | - | 3.35 (0.84–13.4) P=0.088 | - |
| | TT | 1.34 (0.68–2.66) P=0.402 | - | 1.25 (0.53–2.94) P=0.615 | - | 1.85 (0.67–5.07) P=0.234 | - | 0.48 (0.11–2.13) P=0.331 | - |
| rs11169552 | CT/TT | 0.73 (0.42–1.27) P=0.268 | - | 0.77 (0.39–1.53) P=0.457 | - | 0.61 (0.26–1.47) P=0.275 | - | 0.58 (0.15–2.26) P=0.436 | - |
| | TT | 1.11 (0.36–3.41) P=0.855 | - | 0.60 (0.13–2.84) P=0.523 | - | 1.21 (0.21–7.01) P=0.828 | - | 0.87 (0.08–9.05) P=0.906 | - |
| rs7136702 | CT/TT | 1.08 (0.61–1.93) P=0.784 | - | 0.99 (0.49–2.00) P=0.984 | - | 1.31 (0.54–3.17) P=0.546 | - | 0.88 (0.23–3.44) P=0.853 | - |
| | TT | 1.66 (0.74–3.76) P=0.220 | - | 2.46 (0.97–6.24) P=0.058 | - | 1.71 (0.52–5.63) P=0.378 | - | 0.19 (0.03–1.03) P=0.055 | - |
| rs6691170 | TG/TT | 1.22 (0.69–2.16) P=0.500 | - | 0.92 (0.46–1.85) P=0.817 | - | 1.21 (0.49–2.99) P=0.686 | - | 0.65 (0.12–3.47) P=0.616 | - |
| | TT | 1.33 (0.59–2.99) P=0.492 | - | 0.81 (0.29–2.26) P=0.693 | - | 1.05 (0.31–3.56) P=0.943 | - | 0.26 (0.06–1.22) P=0.088 | - |
| rs6687758 | AG/GG | 0.99 (0.57–1.74) P=0.979 | - | 1.03 (0.51–2.09) P=0.932 | - | 1.07 (0.44–2.60) P=0.877 | - | 0.52 (0.13–2.10) P=0.359 | - |
| | GG | 1.93 (0.53–7.00) P=0.320 | - | 1.13 (0.26–5.05) P=0.868 | - | No pts. | - | 0.32 (0.03–3.99) P=0.376 | - |
| rs10936599 | TC/TT | 0.97 (0.56–1.68) P=0.903 | - | 0.98 (0.50–1.92) P=0.952 | - | 1.51 (0.64–3.55) P=0.348 | - | 2.65 (0.61–11.6) P=0.194 | - |
| | TT | 0.76 (0.24–2.40) P=0.643 | - | 5.29 (0.57–49.2) P=0.143 | - | 0.73 (0.05–9.69) P=0.808 | - | No pts. | - |
| rs4925386 | TC/TT | 1.31 (0.76–2.24) P=0.329 | - | 1.08 (0.56–2.07) P=0.819 | - | 0.7 (0.31–1.57) P=0.383 | - | 0.94 (0.25–3.59) P=0.927 | - |
| | TT | 0.90 (0.36–2.22) P=0.820 | - | 0.53 (0.18–1.57) P=0.252 | - | 1.39 (0.36–5.31) P=0.63 | - | 0.75 (0.12–4.63) P=0.758 | - |

| | | Response | | GI toxicity | | Haematological toxicity | | Skin toxicity | |
|---|---|---|---|---|---|---|---|---|---|
| | | TP | VP* | TP | VP* | TP | VP* | TP | VP* |
| rs4444235 | TC/CC | 1.17 (0.64–2.15) P=0.603 | - | 1.16 (0.55–2.46) P=0.701 | - | 0.96 (0.38–2.44) P=0.939 | - | 1.14 (0.25–5.17) P=0.864 | - |
| | CC | 1.8 (0.95–3.41) P=0.070 | - | 0.74 (0.34–1.61) P=0.444 | - | 1.01 (0.35–2.9) P=0.99 | - | 0.47 (0.11–1.93) P=0.292 | - |
| rs9929218 | AG/AA | 0.89 (0.52–1.52) P=0.674 | - | 1.16 (0.61–2.23) P=0.647 | - | 0.41 (0.18–0.94) P=0.035 | 1.29 (0.59–2.82) P=0.521 | 0.44 (0.11–1.72) P=0.236 | - |
| | AA | 3.77 (1.38–10.3) P=0.010 | 3.13 (1.04–9.42) P=0.042 | 1.13 (0.38–3.36) P=0.819 | - | 1.09 (0.26–4.63) P=0.905 | - | 0.14 (0.03–0.67) P=0.014 | 0.07 (0.01–0.55) P=0.011 |
| rs10411210 | TC/TT | 1.10 (0.53–2.27) P=0.793 | - | 0.84 (0.34–2.1) P=0.714 | - | 0.38 (0.13–1.1) P=0.074 | - | 0.78 (0.14–4.32) P=0.775 | - |
| | TT | No pts. | - | 0.31 (0.01–7.62) P=0.477 | - | No pts. | - | No pts. | - |
| rs961253 | AC/AA | 1.20 (0.68–2.12) P=0.533 | - | 0.75 (0.38–1.49) P=0.414 | - | 1.22 (0.52–2.90) P=0.646 | - | 0.31 (0.06–1.64) P=0.168 | - |
| | AA | 0.35 (0.16–0.78) P=0.010 | 1.00 (0.45–2.23) P=0.993 | 2.03 (0.71–5.83) P=0.187 | - | 8.32 (1.48–46.9) P=0.016 | 0.45 (0.12–1.68) P=0.237 | 0.52 (0.09–3.12) P=0.474 | - |
| rs10795668 | AG/AA | 1.41 (0.81–2.48) P=0.225 | - | 0.58 (0.29–1.16) P=0.122 | - | 0.75 (0.31–1.8) P=0.514 | - | 0.61 (0.14–2.68) P=0.513 | - |
| | AA | 2.20 (0.78–6.21) P=0.136 | - | 1.35 (0.48–3.81) P=0.573 | - | 2.4 (0.61–9.48) P=0.213 | - | No pts. | - |
| rs3802842 | AC/CC | 0.89 (0.51–1.53) P=0.667 | - | 0.93 (0.47–1.81) P=0.824 | - | 0.77 (0.32–1.83) P=0.550 | - | 1.83 (0.47–7.08) P=0.38 | - |
| | CC | 0.72 (0.29–1.76) P=0.47 | - | 1.88 (0.62–5.72) P=0.264 | - | 0.25 (0.04–1.65) P=0.150 | - | No pts. | - |
| rs4813802 | GT/GG | 1.71 (0.93–3.11) P=0.082 | - | 0.95 (0.45–2.02) P=0.899 | - | 2.07 (0.81–5.34) P=0.130 | - | 0.19 (0.02–1.62) P=0.128 | - |
| | GG | 1.15 (0.59–2.25) P=0.688 | - | 1.33 (0.58–3.07) P=0.503 | - | 3.27 (1.00–10.7) P=0.050 | 2.14 (0.71–6.47) P=0.179 | 0.22 (0.05–0.92) P=0.038 | 0.88 (0.16-4.91) P=0.884 |

| | | Response | | GI toxicity | | Haematological toxicity | | Skin toxicity | |
|---|---|---|---|---|---|---|---|---|---|
| | | TP | VP* | TP | VP* | TP | VP* | TP | VP* |
| rs1957636 | GA/GG | 0.85 (0.48–1.48) P=0.557 | - | 0.93 (0.47–1.84) P=0.841 | - | 1.08 (0.46–2.56) P=0.858 | - | 0.87 (0.22–3.39) P=0.844 | - |
| | GG | 1.39 (0.62–3.12) P=0.421 | - | 1.44 (0.58–3.61) P=0.435 | - | 2.27 (0.67–7.64) P=0.186 | - | 0.22 (0.03–1.63) P=0.139 | - |
| rs16969681 | TC/TT | 1.00 (0.48–2.09) P=0.989 | - | 2.19 (0.93–5.13) P=0.071 | - | 1.00 (0.32–3.11) P=0.995 | - | 1.87 (0.34–10.4) P=0.472 | - |
| | TT | 0.26 (0.02–3.92) P=0.33 | - | 1.43 (0.06–35.2) P=0.826 | - | No pts. | - | No pts. | - |
| rs11632715 | GA/GG | 0.56 (0.30–1.06) P=0.074 | - | 1.62 (0.76–3.46) P=0.212 | - | 0.65 (0.24–1.70) P=0.376 | - | 5.30 (1.29–21.8) P=0.021 | No pts. |
| | GG | 0.75 (0.41–1.40) P=0.372 | - | 0.52 (0.23–1.20) P=0.125 | - | 0.97 (0.38–2.50) P=0.951 | - | 2.81 (0.53–15.0) P=0.226 | - |

**Table 5.5** - Predictive effects of CRC-risk SNPs on response to and side effects from 12 weeks of cetuximab treatment. Numbers represent HRs with 95% CI in parenthesese. Hazard ratios are "interaction HRs", and should be interpreted as the ratio of the HR (genotype group indicated versus all other patients with a result for that SNP) among patients receiving cetuximab to the HR among patients not receiving cetuximab. TP - Training phase, VP – Validation phase. Significant results in both training and validation phases are shaded. No pts. – no patients in that phase with that genotype. * Though data is not shown, correction for multiple testing was applied after the VP. All observations were non-significant after this correction, barring rs9929218 and its predictive effect on skin toxicity (P=0.022).

| | | | Training phase | | | | Validation phase | | | | Combined | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | n | OR (95% CI) | P | P inter-action | n | OR (95% CI) | P | P inter-action | n | OR (95% CI) | P | P inter-action | $r^2$ | RE P-value |
| 12-week response | AG/AA | - cetuximab | 279 | 1.10 (0.79 – 1.53) | 0.564 | 0.688 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| | | + cetuximab | 170 | 0.98 (0.64 – 1.50) | 0.913 | | -- | -- | -- | | -- | -- | -- | | | |
| | AA | - cetuximab | 50 | 0.49 (0.27 – 0.89) | 0.019 | 0.010 | 40 | 0.47 (0.24 – 0.91) | 0.026 | 0.028 | 90 | 0.48 (0.31 – 0.75) | 0.001 | 0.001 | 0% | 0.001 |
| | | + cetuximab | 31 | 1.86 (0.82 – 4.21) | 0.135 | | 24 | 1.51 (0.62 – 3.64) | 0.361 | | 55 | 1.69 (0.93 – 3.07) | 0.086 | | | |
| Skin toxicity (including HFS) | AG/AA | - cetuximab | 306 | 1.55 (0.43 - 5.57) | 0.500 | 0.693 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| | | + cetuximab | 111 | 1.12 (0.44 – 2.87) | 0.811 | | -- | -- | -- | | -- | -- | -- | | | |
| | AA | - cetuximab | 62 | 7.59 (2.23 – 25.9) | 0.001 | 0.014 | 44 | 3.35 (0.88 – 12.7) | 0.076 | 0.011 | 106 | 5.22 (2.12 – 12.9) | <0.001 | <0.001 | 0% | <0.001 |
| | | + cetuximab | 22 | 1.10 (0.42 – 2.87) | 0.851 | | 19 | 0.26 (0.06 – 1.15) | 0.075 | | 41 | 0.72 (0.32 – 1.61) | 0.42 | | | |
| 12-week HFS | AG/AA | - cetuximab | 311 | 1.55 (0.43 – 5.57) | 0.500 | 0.693 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| | | + cetuximab | 144 | 1.12 (0.44 – 2.87) | 0.811 | | -- | -- | -- | | -- | -- | -- | | | |
| | AA | - cetuximab | 63 | 6.04 (1.65 – 22.2) | 0.007 | 0.225 | 44 | 3.55 (0.92 – 13.7) | 0.065 | -- | 107 | 4.68 (1.84 – 11.9) | 0.001 | 0.030 | -- | -- |
| | | + cetuximab | 27 | 1.91 (0.51 – 7.10) | 0.334 | | -- | -- | -- | | 58 | 0.84 (0.25 – 2.83) | 0.778 | | | |

**Table 5.6** - Predictive analyses of rs9929218 in E-cadherin on response to and side effects from 12 weeks of cetuximab treatment. Odds ratios should be interpreted as the odds in the genotype group indicated versus all other patients with a result for that SNP.

149

We investigated whether the somatic mutation status of the CRCs influenced the predictive effect on response to cetuximab. Patients homozygous for the rs9929218 minor allele in the germline and with *KRAS, BRAF* or *NRAS* mutant tumours showed a significantly improved response to cetuximab (OR 1.69, 95% CI 0.61-4.74 with cetuximab; OR 0.25, 95% CI 0.11-0.55 without cetuximab; P for interaction 0.004). In contrast, patients homozygous for the rs9929218 minor allele and with *KRAS, BRAF* and *NRAS* wild-type tumours did not show a statistically significant response (OR 1.51, 95% CI 0.48-4.70 with cetuximab; OR 0.85, 95% CI 0.40-1.83 without cetuximab; P for interaction 0.35).

### 5.3.5 Delineating the mechanism underlying the response to cetuximab

rs9929218 lies within intron-2 of the *CDH1* gene encoding E-cadherin, in strong LD with rs16260 (Pittman *et al.* 2009), also called -160C>A, in the CDH1 promoter (Figure 5.3). rs16260 has been shown to downregulate CDH1 expression (Li *et al.* 2000). Therefore, we genotyped rs16260 in patients from COIN and COIN-B and observed strong LD with rs9929218 ($r^2$=0.915, D'=0.984). We found similar prognostic effects for rs16260, where patients homozygous for the minor allele (AA) (equating to ~7% of patients) showed a significant decline in OS as compared to patients with the CA or CC genotypes (HR 1.38, 95% CI 1.15-1.67, P=0.001), equating to a median decrease in life expectancy of 3.9 months. Patients homozygous for the rs16260 minor allele had a poor prognosis for 12-week response to fluoropyrimidine and oxaliplatin-based treatment as compared to patients that were heterozygous or homozygous wild-type (OR 0.58, 95% CI 0.37–0.93, P=0.022), and had an increased risk of skin toxicity (OR 5.29, 95% CI 2.11–13.3, P<0.001), specifically HFS (OR 4.78, 95% CI 1.80–12.7, P=0.002) (Table 5.7).

rs16260 also predicted response to cetuximab (Table 5.8). Patients homozygous for the rs16260 minor allele showed significantly improved 12-week response (OR 1.70, 95% CI 0.93–3.11 with cetuximab, OR 0.58, 95% CI 0.37–0.93 without cetuximab; P for interaction=0.007) and this was most apparent in patients with *KRAS, BRAF* or *NRAS* mutant tumours (OR 1.47, 95% CI 0.53–4.10 with cetuximab and OR 0.22, 95% CI 0.09–0.51 without cetuximab; P for interaction=0.005).
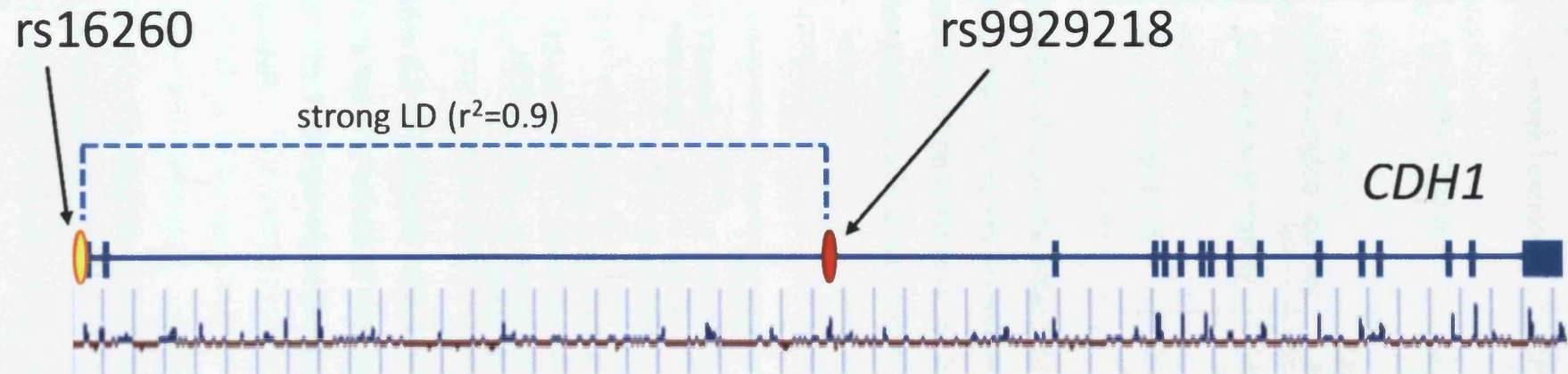
**Figure 5.3** – Schematic representation of the *CDH1* gene structure with exons shown by vertical blue lines. SNPs rs9929218 and rs16260 are indicated. Conservation across the gene (according to USCS Genome Bioinformatics database) is shown beneath the schematic, with blue peaks representing conserved islands. As expected, conserved islands align well with expressed sequence, and rs9929218. Other conserved islands are present within the second intron of the gene, though none coincide with SNPs in high LD ($r^2$>0.8) with rs9929218.

|  | | n | OR (95%CI) | P-value |
|---|---|---|---|---|
| 12-week response | CA/AA | 524 | 1.09 (0.86 – 1.38) | 0.482 |
| | AA | 79 | 0.58 (0.37 – 0.93) | 0.022 |
| 12-week GI toxicity | CA/AA | 464 | 1.33 (0.97 – 1.82) | 0.077 |
| | AA | 74 | 0.95 (0.51 – 1.74) | 0.858 |
| 12-week haematological toxicity | CA/AA | 528 | 1.12 (0.77 – 1.63) | 0.552 |
| | AA | 89 | 1.15 (0.57 – 2.32) | 0.691 |
| 12-week skin toxicity | CA/AA | 576 | 0.93 (0.41 – 2.10) | 0.868 |
| | AA | 93 | 5.29 (2.11 – 13.3) | 0.001 |
| 12 week HFS | CA/AA | 584 | 0.92 (0.40 – 2.16) | 0.855 |
| | AA | 94 | 4.78 (1.80 – 12.7) | 0.002 |

**Table 5.7** - Prognostic effect of rs16260 in E-cadherin on response to and side effects from 12 weeks of oxaliplatin and fluoropyrimidine based chemotherapy. Odds ratios should be interpreted as the odds in the genotype group indicated versus all other patients with a result for that SNP.

| | | | n | OR (95% CI) | P-value | P interaction |
|---|---|---|---|---|---|---|
| 12-week response | CA/AA | - cetuximab | 524 | 1.09 (0.86 – 1.38) | 0.482 | 0.891 |
| | | + cetuximab | 328 | 1.06 (0.79 – 1.43) | 0.695 | |
| | AA | - cetuximab | 79 | 0.58 (0.37 – 0.93) | 0.022 | 0.007 |
| | | + cetuximab | 54 | 1.70 (0.93 – 3.11) | 0.087 | |
| 12-week HFS | CA/AA | - cetuximab | 584 | 0.92 (0.40 – 2.16) | 0.855 | 0.941 |
| | | + cetuximab | 316 | 0.96 (0.51 – 1.84) | 0.911 | |
| | AA | - cetuximab | 94 | 4.52 (1.71 – 11.9) | 0.002 | 0.040 |
| | | + cetuximab | 55 | 0.89 (0.26 – 3.01) | 0.854 | |

**Table 5.8** - Predictive analyses of rs16260 in E-cadherin on response to and side effects from 12 weeks of cetuximab treatment. Odds ratios should be interpreted as the odds in the genotype group indicated versus all other patients with a result for that SNP.

## 5.4 Discussion

Here, we have shown that multiple low penetrance CRC-risk alleles affect survival in patients with aCRC, highlighting the relationship between genetic variants acting as both predisposition and prognostic alleles. Tenesa and colleagues (2010) previously analysed a population based cohort of 2,838 Scottish CRC patients but did not find any GWAS CRC-risk SNPs (including rs16892766, rs9929218, rs10411210 and rs10795668) associated with survival. However, their cohort primarily consisted of cases with Stage 1-3 disease and therefore had fewer cases that were very ill at presentation and died soon afterwards, so are likely to have had less power to identify the findings described herein. However, it must also be considered that some of the survival effects observed here might actually be the result of altered response to oxaliplatin and fluoropyrimidine based chemotherapy since we did not validate the prognostic effects across all arms of the trial. Furthermore, it must be considered that if multiple testing is accounted for, only a minority of the observations made here remain significant; specifically the prognostic effect on OS of rs9929218 and the prediction of improved skin toxicity again in homozygous carriers of the rs9929218 minor allele. Though a number of observations are removed, it is still clear that rs9929218 in E-cadherin plays a critical role in defining prognosis and response in aCRC

E-cadherin functions as a transmembrane glycoprotein that is critical in the establishment and maintenance of intercellular adhesion, cell polarity and tissue morphology and regeneration (Takeichi, 1991). Inactivation of E-cadherin by promoter hypermethylation is a common event in CRC (Wheeler *et al.* 2001, Garinis *et al.* 2002) and its loss represents the defining feature of the epithelial to mesenchymal transition (EMT) - a change in cellular phenotype that sees a reduction in cell-cell adhesion structures, as well as alterations of polarity and cytoskeletal structure (Klymkowsky and Savagner, 2009). Here, we show that rs9929218 in E-cadherin is an independently validated predictive biomarker of response to cetuximab. Interestingly, other studies have also implicated E-cadherin in altering response to cetuximab. Urothelial carcinoma cells expressing E-cadherin were found to be sensitive to cetuximab whilst those lacking E-cadherin were resistant and knockdown of E-cadherin in the sensitive cells led to reduced response (Black *et al.* 2008). In non-small cell lung cancer cell lines, high levels of E-cadherin expression

correlated with sensitivity to cetuximab (Nikolova *et al.* 2009) and in *KRAS* wild-type epidermoid carcinoma cells cetuximab efficacy was dependent upon E-cadherin expression (Oliveras-Ferraros *et al.* 2010). However, the exact mechanism through which rs9929218 influences this response remains unclear. This is discussed further in chapter 7.

### 5.4.1 19q13 and 10p14

We found that two SNPs, rs10411210 at 19q13 and rs10795668 at 10p14, affected OS. Compared to the 8q23 (rs16892766) and 16q22 (rs9929218) loci, little is known about the mechanisms of risk at these loci. rs10411210 maps to an LD block that encompasses the Rho GTPase binding protein 2 gene (*RHPN2*). This encodes a Rho GTPase involved in regulation of the actin cytoskeleton and cell motility (Peck *et al.* 2002). RhoA proteins have previously been associated with CRC risk and prognosis (Bellovin *et al.* 2006). It is perhaps interesting that *RHPN2*, along with *CDH1*, encode proteins with overlapping functions in regulating the EMT process. In-turn, our data points to a role for EMT in defining the course of the disease and patient prognosis, as well as to response to therapy.

rs10795668 maps to an 82kb block of LD at the 10p14 locus. Of note, there are no known protein coding transcripts within this region and the nearest predicted genes are located 0.4Mb proximal (BC031880) and 0.7Mb distal (LOC389936). As such, it will likely be very difficult to identify the causal mechanism at this locus. However, studies at 8q24 (Wright *et al.* 2010) prove that SNPs can have mechanistic effects across considerable distances.

### 5.4.2 Future work

Four independent loci have been found to alter survival in patients with aCRC. It will be important to determine what role these variants play in patients with earlier stages of CRC. It is likely that this will only be possible through the use of sufficiently large cohorts with large follow-up times. Further to this, and relating to the role these variants play in defining risk, it will be important to determine the mechanisms by which these variants confer their effects through the use of functional experiments (including expression analyses). Once a greater understanding of these mechanisms is obtained, it might be possible to target them for therapy.

154

# Chapter Six - Identifying biallelic high penetrance CRC susceptibility alleles in the BER, MMR and OxDR pathways

## 6.1 Introduction

### 6.1.1 *MUTYH* and CRC predisposition

It has previously been shown that the pattern of somatic mutagenesis in colorectal tumours can be used to identify novel CRC predisposition genes (Al-Tassan *et al.* 2002, reviewed in Cheadle and Sampson 2003). Investigators studied a British family (Family N) with three affected siblings with multiple CRAs and carcinoma that did not carry an inherited *APC* gene defect. To provide a clue as to the underlying genetic defect, the entire *APC* open reading frame was sequenced in eleven tumours from Family N, revealing eighteen somatic *APC* mutations, 15 of which were G:C→T:A transversions. This class of mutations accounted for only ~10% of reported somatic *APC* mutations, with frameshift mutations and LOH being the more usual classes of mutations leading to somatic inactivation of *APC* in colorectal tumours. Comparison of the findings in Family N with a database of somatic *APC* mutations from sporadic and FAP-associated colorectal tumours, confirmed that the excess of G:C→T:A transversions in Family N ('the mutator phenotype') was highly significant ($P=10^{-12}$). Previous studies in bacteria and yeast had shown a mutator phenotype with this characteristic in *MutY* and *Ogg1*-deficient strains and subsequent studies of the corresponding human orthologues showed that patients from Family N were germline compound heterozygotes for the nonsynonymous variants Y179C and G396D (formerly Y165C and G382D) in the human orthologue of *MutY* (termed *MUTYH*). Further studies (Jones *et al.* 2002, Sampson *et al.* 2003) have confirmed that biallelic mutations in *MUTYH* predispose to colorectal tumours and account for around a quarter of *APC* mutation negative patients with over 10 CRAs.

Together the two MUTYH nonsynonymous mutations Y179C and G396D (that have minor allele frequencies (MAFs) of ~0.5% and 1%, respectively, in European normal controls) account for over 80% of mutant alleles in the British population (Sampson *et al.* 2003; Cheadle and Sampson, 2007). Significantly, both of these variants were documented in the public domain database dbSNP prior to the discovery that they

155

were colorectal tumour predisposition alleles suggesting that this database may be a rich source of other, as yet undiscovered, CRC predisposition alleles. MUTYH functions as a BER DNA glycosylase and functionally cooperates with the MMR system to recognise and repair oxidative DNA damage as described in chapter 1.

## 6.1.2 Aims

Here, we screened our collection of 2,186 aCRC and 2,176 control samples for all nonsynonymous SNPs annotated within dbSNP from every gene in the BER, MMR and oxidative damage repair pathways (OxDR) pathways to identify other autosomal recessive 'MUTYH-like' genes. We searched for biallelic combinations of high penetrance alleles within each gene that were only found in cases with CRAs or CRCs and never found together in controls

## 6.2 Materials and Methods

### 6.2.1 Samples

*(i) Cases with advanced CRC and controls*

All 2,186 aCRC cases and 2,176 UKBS healthy controls have previously been described (chapters 3, 4 and 5)

*(ii) Cases with multiple CRAs and controls*

We analysed 163 unrelated index cases with multiple CRAs with or without carcinoma that were recruited from either UK or Dutch polyposis registers. No cases carried biallelic MUTYH mutations, or a truncating mutation in APC. We also analysed DNA samples from a panel of 480 unrelated healthy Caucasian controls purchased from the Health Protection Agency (Salisbury, UK).

*(iii) POPGEN cohort and controls*

The POPGEN cohort consisted of 2,147 histologically proven colorectal carcinoma patients. Patients were operated on between 2002 and 2005 in surgical departments in northern Germany. HNPCC patients were excluded. The patient cohort had a median age at diagnosis of 63 years (ranging from 18 to 92 years) and consisted of 45% females. 2,787 geographically matched control samples from northern Germany were also used.

## 6.2.2 Power considerations

We searched for biallelic combinations of rare nonsynonymous, truncating or splice-site mutations within each gene that were only found in cases with CRAs or CRCs and never found together in controls. Such an approach would have led to the identification of MUTYH as a CRC predisposition gene as exemplified by the study of Farrington *et al.* (2005) who studied 2,239 cases with CRC and 1,845 controls. They found 11 cases harboured biallelic Y179C/G396D mutations whereas such combinations were never found in controls. With the sample size of 2,200 cases and 2,200 controls we had >95% power to detect the previously observed difference between the proportions of cases and controls harbouring biallelic MUTYH mutations at a 0.05 significance level.

## 6.2.3 Genes/Variants chosen for analysis

Genes involved (or implicated) in the BER, MMR and OxDR pathways were selected from the comprehensive Wood's list of repair genes (Woods *et al.* 2005) and through literature searches. Variants within these genes were searched for using dbSNP (build version 129) (http://www.ncbi.nlm.nih.gov/projects/SNP/) or through literature review. The criterion for inclusion in this project was nonsynonymous SNPs in BER, MMR and OxDR pathway genes with a frequency of up to 2%, or with no frequency data. SNPs chosen from the respective pathways are presented in Tables 6.1, 6.2 and 6.3.

## 6.2.4 Exclusion criteria

Given the divergence of SNP frequencies between populations, we excluded known non-Caucasian samples from our analysis - these were identified by reviewing the original patient notes. In addition, genes previously associated with CRC were excluded from the analysis since we were targeting novel tumorigenic mechanisms. The only exception to this was *MUTYH*, in which we chose Y179C and G396D to act as positive controls.

| Gene Name (aka) | Function | Evidence for involvement in BER | SNPs successfully analysed | SNPs that failed locus conversion or failed genotyping | References |
|---|---|---|---|---|---|
| APEX1 (APE1, REF1) | Human endonuclease. Incises DNA backbone at AP site | Purified human APEX1 is able to remove 3'-8oxoG lesions | No nonsynonymous SNPs (nsSNPs) | rs34632023, rs2307486, rs1803118, rs33956927 | Parson et al. 2005 |
| APEX2 | Human endonuclease. Incises DNA backbone at AP site | APEX2 deficient mice are smaller than wild-type mice. Implicated in post-replicative BER | No nsSNPs | rs11541055, rs2301416, rs3180303, rs3180304 | Ide et al. 2003 |
| APTX | Catalyses the release of adenylate groups covalently linked to 5' phosphate termini | Reduced observation of DNA nicks and increased ligation products through addition of Aptx in E.coli | L74M | rs12001066 | Rass et al. 2007 |
| ERCC6 (CSB) | Principal function in transcription coupled repair. May influence repair of oxidised lesions through interaction with lesion-specific DNA glycosylases | Deficient cells have reduced capacity to repair 8oxoG | No nsSNPs | rs2228525, rs2229761, rs34704611, rs34917815, rs41281957, rs4253046, rs4253206, rs4253208, rs4253219, rs4253227, rs4253228, rs4253229, rs4253230 | Dianov et al. 1999 |
| FEN1 | Endonuclease that functions in long patch BER. Processes Okazaki fragments | FEN1 deficient cells display hypersensitivity to $H_2O_2$, suggesting defective BER | rs7931165, rs4989588, rs11541090 | rs1803573, rs4989587, rs4989586 | Asagoshi et al. 2010 |
| LIG1 | Joins DNA fragments generated during DNA replication and repair | Extracts from cell lines defective in the LIG1/PCNA interaction are defective in long patch BER | V753M, rs3730947, rs3730980 | rs11668325, rs11879148, rs12981963, rs34087182, rs3730855, rs3730863, rs3730911, rs3730933, rs3730966, rs3731003, rs3731008, rs4987070, rs4987181 | Levin et al. 2000 |
| LIG3 | Forms a complex with XRCC1. Seals nicks formed during BER | Extracts from LIG3 deficient cell lines undergo defective BER | rs1076992, rs36021499, rs1802880, rs4986974, rs35956074 | rs35501315, rs4986973 | Tomkinson et al. 2001 |
| MBD4 | Mono-functional DNA glycosylase that removes thymine from T:G mispairs | Mbd4 mutant mice show a 2- to 3-fold increase in C>T transitions | No nsSNPs | rs2307289, rs2307293, rs2307296, rs2307298 | Wong et al. 2002 |
| MPG (AAG) | Mono-functional DNA glycosylase. | Cells deficient in MPG are | No nsSNPs | E240D, rs25671, rs2234949, rs2266607, | Engelward et al. 1997 |

| | | | | | |
|---|---|---|---|---|---|
| | Excision of 3- and 7-methyladenine | hypersensitive to killing by alkylating agents | | rs2308312, rs2308315, rs3176383, rs769193 | |
| MTH1 | Hydrolyses 8-oxo-dGTP, removing it from the nucleotide pool | Knock-out mice displayed greater numbers of lung, liver and stomach tumours than controls | K155I, V83M | rs11547459, rs57000894 | Tsuzuki et al. 2001 |
| MUTYH | Bi-functional DNA glycosylase that excises mispaired adenine opposite 8oxoG | Defects in MUTYH contribute to defective BER and a subsequent increase in G>T transversions | rs36053993 (G396D), rs34612342 (Y179C) | No nsSNPs | Al-Tassan et al. 2002 |
| NEIL1 | Bi-functional DNA glycosylase. Suggested to act as a 'back-up' system to OGG1 and NTH | Deficient murine ES cells are more sensitive to the cytotoxic effects of γ-irradiation than controls | NEIL1_434+2T>C, R339Q, rs5745905, rs7183491 | P208S, rs5745906, rs5745907 | Morland et al. 2002; Rosenquist et al. 2003; Fortini et al. 2003 |
| NEIL2 | Bi-functional DNA glycosylase. Active towards several cytosine-derived lesions | Activity of NEIL2 increases following oxidative stress | R164T, rs8191611, rs8191664, rs8191666 | rs35400293, rs8191612, rs8191613, rs57683451 | Das et al. 2007 |
| NEIL3 | Bi-functional DNA glycosylase that recognises specific oxidized purines (but not 8oxoG) | Expression of its homologue in E.coli lacking other glycosylases reduced spontaneous mutation frequency | R315Q, rs34112288, rs17064676, rs36005630 | D132V, rs35418725, rs61748603, rs58603054, rs58437627, rs58297365 | Liu et al. 2010 |
| NTHL1 (NTH1) | Bi-functional DNA glycosylase that can remove oxidized pyrimidines | Shown to excise multiple oxidized pyrimidines | Q90X, rs2233522, rs1805378 | rs3087468, rs3211977, rs3087469 | Aspinwall et al. 1997 |
| OGG1 | Bi-functional DNA glycosylase. Excises 8-oxo-G from DNA | Cell extracts containing OGG1 display lyase activity towards the 8oxoG lesions. Introduction of OGG1 to a spontaneous mutator strain of E.coli rescues them - reducing the rate of mutation | R46Q, A153T, G308E, I321T | rs1052140, rs11548133, rs1801128, rs1805373, rs3219012, rs3219014, rs56053615 | Aburatani et al. 1997; Radicella et al. 1997 |
| PARP1 | Binds with high affinity to double- and single-strand | Knockout cells are sensitive towards alkylating agents | No nsSNPs | rs1059011, rs1059040, rs1136420, rs1805409, rs3219057, rs3219145, | Allinson et al. 2003; Ménessier-de |

159

| | | | | rs3738708 | Murcia et al. 1997; Wang et al. 1997; Vodenicharov et al. 2000; Wood et al. 2005 |
|---|---|---|---|---|---|
| *PARP2* | Binds with high affinity to double- and single-strand DNA breaks. Acts to protect strand interruptions | Deficient cells have delayed DNA strand break re-sealing. Knock-out mice are sensitive to ionising radiation and display genomic instability | rs2275010 | rs1128782, rs3093905, rs3093906, rs3093925 | Schreiber et al. 2002; Ménessier-de Murcia et al. 1997 |
| *PCNA* | Required for the repair initiation step of MMR and repair DNA resynthesis | Extracts from cell lines defective in the LIG1/PCNA interaction are defective in long patch BER | rs1050525 | No nsSNPs | Levin et al. 2000 |
| *PNKP (PNK)* | Kinase that removes 3' phosphate groups and adds 5' phosphate groups to hydroxyl termini | Shown to remove the 3' phosphate from NEIL1 generated termini | No nsSNPs | rs11555414, rs11671530, rs34472250, rs3739168, rs3739173, rs3739185, rs3739186, rs3739206 | Wiederhold et al. 2004 |
| *POLβ* | In short patch BER, extends the 3'-OH terminus and removes the 5' phosphate | Cell lines carrying a homozygous deletion mutation in *POLβ* show defective BER | rs56121607, rs12678588, rs10091081, rs3136797 | No nsSNPs | Sobol et al. 1996 |
| *POLγ* | Localised to mitochondria. Replicates and repairs mitochondrial DNA | Shown to excise deoxyribose phosphate from AP sites | G268A, rs2307437, rs2307447, rs2307450, rs41549716, rs56119329, rs56410699, rs60636456, rs11546842 | rs2307440, rs2307442, rs28567406, rs3087373, rs3176162, rs61472028, rs56047213 | Longley et al. 1998; Pinz and Bogenhagen, 2006 |
| *SMUG1* | Mono-functional DNA glycosylase. Targets uracil and uracil derivatives | Initially isolated through its ability to bind DNA damage lesions. Shown to target uracil residues | rs2233919 | rs1043202, rs3136389, rs3087405, rs2233920 | Haushalter et al. 1999; Matsubara et al. 2004 |
| *TDG* | Mono-functional DNA glycosylase. Targets thymine and uracil when mispaired with guanine | Purified protein excises mispaired thymine in HeLa cells | T19M, I134V, Q391K | rs12367528, rs17853764, rs3953596, rs3953597, rs3953598 | Neddermann and Jiricny, 1993 |
| *TDP1* | Member of the | Shown to efficiently | No nsSNPs | rs34452707, | Interthal et al. |

| | | | | rs35114462, | 2005; Brata |
|---|---|---|---|---|---|
| | phospholipase D super-family that hydrolyses the phosphodiester bond between a DNA 3' end and a tyrosyl moiety. Known to interact with BER proteins | cleave at AP sites using duplex DNA - efficiency similar to that of *NEIL1*. Mitochondrial BER depends on TDP1 activity | | rs35455108, rs35973343 | Das *et al.* 2010 |
| *UNG1* | Mitochondrial. Mono-functional DNA glycosylase. Targets Uracil for excision. | Ung-/- mice exhibit a moderate increase in genome-wide spontaneous mutagenesis | No nsSNPs | No nsSNPs | An *et al.* 2005; Hazra *et al.* 2007 |
| *UNG2* | Nuclear. Repairs deaminated cytosine amongst other lesions. Binds PCNA. | Nuclear extracts of human UNG2 initiates BER of plasmids containing U:A and U:G | No nsSNPs | No nsSNPs | Kavli *et al.* 2002 |
| *XRCC1* | Scaffolding protein involved in repair of single-strand breaks. Has no catalytic activity itself | xrcc1 mutant cells are partially defective in ligation of base excision repair patches. Interacts physically and functionally with OGG1 | rs25474, rs2307184, rs2682557, rs2307167, rs2307186, rs2307191 | rs2307177, rs2307166, rs41561817, rs2271980, rs2307188, rs2307180, rs2307171, rs11553659 | Cappelli *et al.* 1997; Marsin *et al.* 2003 |

**Table 6.1** - Genes implicated in BER. Listed are SNPs (with MAFs ≤2%) successfully analysed and those that either failed *in silico* design or failed genotyping. Though they initially failed genotyping on the illumina platform, Y179C and G396D of *MUTYH* (bold) were successfully genotyped via combined use of restriction digest and Taqman based assays.

| Gene Name (aka) | Function | Evidence for involvement in MMR | SNPs successfully analysed | SNPs that failed locus conversion or failed genotyping | References |
|---|---|---|---|---|---|
| BCL2 | Proto-oncogene that may regulate MSH2/6 complex. Pro-survival factor implicated in a number of cancers. | Deletion of specific domains of Bcl2 influences its interaction with MSH6 and associates with enhanced MMR efficiency | No nsSNPs | rs1800477 | Hou et al. 2007 |
| DNMT1 | Maintains global methylation after DNA replication | Deficient cells exhibit higher MSI than normal cells | rs16999593, rs35622143, DNMT1 | No nsSNPs | Ge Guo et al. 2004 |
| EXO1 | DNA exonuclease. Can physically interact with MMR proteins MSH2, MLH1 and MSH3 | Mice lacking exon 6 of Exo1 exhibited defective MMR | rs4149864, rs4149967 | rs4149865, rs4149962, rs4149964, rs60358723, rs60320391, rs60293068, rs59790129, rs59734256, rs59699975, rs57175702, rs56561721, rs4149966, rs4149978 | Bardwell et al. 2004 |
| HMG1 (HMGB1) | Chromosomal proteins – bind major groove of DNA causing local distortion of the double helix | Protein purified through its ability to repair mismatches in depleted HeLa cells | rs2388544 | rs3956321, rs4597168, rs11546137 | Zhang et al. 2008; Yuan et al. 2004 |
| MAX | Transcription factor that associates with MSH2 | May exert effects through its interaction with MSH2. May also help regulate MLH1 and MSH2 expression | rs17852278 | No nsSNPs | Partlin et al. 2003; Bindra and Glazer, 2007 |
| MED1 | Involved in regulating transcription. Interacts with MLH1. Associated with the integrity of the MMR system | Cells lacking MED1 show decreased steady state amounts of MMR proteins (including Msh2, Msh6, Mlh1), possibly through a post-transcriptional mechanism | rs35668211, rs35379518, rs1139823, rs1139821, rs61673563 | rs36022909, rs1139828, rs1139822, rs1139820 | Cortellino et al. 2003; Bellacosa et al. 1999 |
| MRE11 | Nuclear protein involved in a number of maintenance processes. May be | MRE11 deficiency leads to a significant increase in MSI for both mono- and | R488C | rs1061945, rs116679717, rs1805367 | Vo et al. 2005 |

162

| | | | | | |
|---|---|---|---|---|---|
| | involved in 3' directed mismatch repair | dinucleotide sequences in HeLa cells | | | |
| *MSH4* | Part of MMR machinery | No evidence for role in HNPCC. Included based on relationship to other MMR proteins | rs5745329, rs17853142 | rs5745311, rs5745324, rs60345283, rs60751948 | |
| *MSH5* | Part of MMR machinery | No evidence for role in HNPCC. Included based on relationship to other MMR proteins | rs1802127, rs28381349, rs28399977, rs28381358, rs56200818 | rs28399976, rs45468693, MSH5 | |
| *MYC* | Transcription factor associated with MLH1 | Deregulated expression linked with influence on MMR activity | No nsSNPs | rs114570780, rs61755060, rs4645960, rs4645968 | Partlin *et al.* 2003 |
| *PMS2L3 (PMS5, PMSR3)* | MutL homolog of unknown function | Mutagenised yeast led to isolation of PMS2L3, along with other known PMS proteins | No nsSNPs | rs3203011, rs2599335, rs2599333, rs2599332, rs794377 | Jeyaprakash *et al.* 1994 |
| *PMS2L4 (PMS6)* | MutL homolog of unknown function | Mutagenised yeast led to isolation of PMS2L4, along with other known PMS proteins | rs7784159 | rs7783101, rs7806476, rs1136533, rs10233400, rs11559285, rs17433617, rs58072630, rs12531701 | Jeyaprakash *et al.* 1994 |
| *POLδ1* | Directly involved in DNA replication – specifically lagging strand synthesis. Intrinsic proof reading exonucleases cooperate with MMR | Depleted HeLa cells had MMR activity restored by addition of calf thymus DNA polymerase delta | rs3218775, rs1052471, rs1726803, rs3219457, rs41563714, rs8105725, rs2230243 | rs9282830, rs55955638, rs3218773, rs9282831, rs2230241, rs58282823, rs3218750, rs11550555, rs3218772, rs41554817 | Longley *et al.* 1997 |
| *POLε* | Directly involved in DNA replication – specifically leading strand synthesis. Intrinsic proof reading exonucleases cooperate with MMR | Like POLδ1, mutant mice display increased rate of mutation and spontaneous cancers | rs11146982, rs5744758, rs5744777, rs5744822, rs5744834, rs5744845, rs5744888, rs12315832, rs5744739, rs5744760, rs5744947, rs5744950, rs5744991, rs7487595, rs5744955, rs5744971, rs41561818, rs5744800, rs56102033 | rs5745067, rs5744935, rs58151768, rs5744933, rs36120395, rs5744752, rs60851699, rs57754437, rs57061760, rs56903770, rs56081968, rs11146986, rs5744972, rs5744943, rs5744856, rs5744801, rs5744734, rs5744799, rs5744904, rs5744948, rs5745021, rs5745068, rs58916399 | Albertson *et al.* 2009 |

163

| Gene | Function | Evidence | Successful SNPs | Failed SNPs | Reference |
|---|---|---|---|---|---|
| *RECQL* (*RECQ1*) | DNA helicase. Interacts with MMR factors | Stimulates the MSH2/6 complex | rs1065751, rs2230003, rs3207640 | rs6499, rs17849407, rs4987215, rs6501, rs3983531 | Doherty *et al.* 2005 |
| *RFC* (*RFC1*) | Multi-subunit ATPase that functions in DNA replication and repair. Required for PCNA loading | A mutant allele of RFC1 associates with a 10-fold increase in number of dinucleotide repeat instability | rs12502450, rs17288828, rs17335452, rs2066791, rs2306598, rs28903095, rs1057749, rs1057751, rs17854711, rs55704262, rs55734630 | rs28903096, rs1135544, rs1057747, rs56273953, rs17419994, rs34586398 | Masih *et al.* 2008; Xie *et al.* 1999 |
| *RFX* (*RFX1*) | Transcription factor. Stimulates 5' to 3' mismatch provoked excision in vitro | RFX knock-out in HeLa cells creates MSI phenotype | rs2305780 | rs60197642 | Zhang *et al.* 2008 |
| *RPA* (*RPA1*) | Stimulates excision, stabilises the ensuing gap against endonuclease attack and promotes repair DNA synthesis | Wild-type, but not mutant, RPA rescues MMR in RPA-depleted cell lysate | rs12449332, rs41547814 | rs56805137, rs55800538 | Lin *et al.* 1998 |

**Table 6.2** - Genes implicated in MMR. Listed are SNPs (with MAFs $\leq 2\%$) successfully analysed and those that either failed *in silico* design or failed genotyping. There was a small overlap between the BER and MMR pathways. The following genes from the BER pathway have also been implicated in MMR but were not included here; *FEN1, LIG1, MBD4, PCNA.*

| Gene Name (aka) | Function | Evidence for involvement in OxDR | SNPs successfully analysed | SNPs that failed locus conversion or failed genotyping | References |
|---|---|---|---|---|---|
| JWA | Reacts as part of cellular response to environmental exposures such as heat and chemical stress | JWA upregulated in oxidative stress cell culture models. | rs6803501, rs11542218 | rs11542221, rs1802400 | Chen et al. 2007 |
| POLL | DNA polymerase that functions in non-homologous end joining and DSBR | Fibroblasts from POLL null mice are hypersensitive to oxidative DNA damaging agents | rs55727130 | rs55978126, rs3209099 | Braithwaite et al. 2005 |
| TIGAR | p53 inducible protein that regulates glycolysis and protects against oxidative stress | Expression levels influence intracellular ROS levels | No nsSNPs | No nsSNPs | Bensaad et al. 2006 |

**Table 6.3** - Genes implicated in OxDR. Listed are SNPs (with MAFs $\leq 2\%$) successfully analysed and those that either failed *in silico* design or failed genotyping. There was significant overlap between OxDR, BER and MMR pathways. The following genes from the BER and MMR pathways have also been implicated in general OxDR but were not included here; *APEX1, FEN1, HMG1, LIG3, MTH1, MUTYH, NEIL1, NEIL2, NEIL3, NTHL1, OGG1, PCNA, POLβ, POLδ1, RFC, SMUG1, UNG1, XRCC1.*

### 6.2.5 Genotyping

Genotyping was performed by Illumina's Fast-Track Genotyping Services (San Diego, CA) using their high throughput BeadArray™ technology.

Since genotyping of the Y179C and G396D control variants failed on the illumina chip we used a combination of restriction digest and Taqman-based assays. For restriction digests, PCR amplification was performed in 25μl reaction volumes using standard reaction conditions. For Y179C we amplified exon 7 and carried out restriction digests using Mwol (NEB). For G396D we amplified exon 13 and carried out restriction digests using BglII (Pharmacia). Restriction digest of OGG1 samples that failed genotyping of G308E on the Illumina platform was carried out using AvaII (NEB).

### 6.2.6 Genotyping of the POPGEN samples

Genotyping of *OGG1* G308E variant was carried out using a custom designed Taqman SNP genotyping assay (Applied Biosystems).

### 6.2.7 Sanger sequencing

We sequenced the ORFs and splice sites of those patients that carried single heterozygous *MUTYH* (Y179C and G396D), *OGG1* (G308E) and *MTH1* (V83M) variants. Primer sequences used for amplification and sequencing are shown in Tables 6.4, 6.5 and 6.6. Sequencing was performed using Applied Biosystems ABI Prism 3100 genetic analyzer. Sequences were analysed using Sequencher software (Sequencher 4.6 - Build 2496).

| Exon | Forward primer sequence | Reverse primer sequence |
|---|---|---|
| 1 | CTTCCCCTCTCCCAGAGC | ATCCCCGACTGCCTGAACC |
| 2 | CTGCATTTGGCTGGGTCTTT | CGCACCTGGCCCTTAGTAAG |
| 3 | GGCCAGAAACTTAGCCACAG | CAACCCCAGATGAGGAGTTAGG |
| 4 | CTCATCTGGGGTTGCATTGA | GGGTTGGCATGAGGACACTG |
| 5 | GGGCAGGTCAGCAGTGTC | TACACCCACCCCAAAGTAGA |
| 6 | TACTTTGGGGTGGGTGTAGA | AAGAGATCACCCGTCAGTCC |
| 7 | GGGACTGACGGGTGATCTCT | TTGGAGTGCAAGACTCAAGATT |
| 8 | CCAGGAGTCTTGGGTGTCTT | AGAGGGGCCAAAGAGTTAGC |
| 9 | AACTCTTTGGCCCCTCTGTG | GAAGGGAACACTGCTGTGAA |
| 10 | GTGCTTCAGGGGTGTCTGC | TGTCATAGGGCAGAGTCACTCC |
| 11 | TAAGGAGTGACTCTGCCCTATG | ACTGGAATGGGGCTTCTGAC |
| 12 | AGCCCCTCTTGGCTTGAGTA | TGCCGATTCCCTCCATTCT |
| 13 | AGGGCAGTGGCATGAGTAAC | GGCTATTCCGCTGCTCACTT |
| 14 | TTGGCTTTTGAGGCTATATCC | CATGTAGGAAACACAAGGAAGTA |
| 15 | TGAAGTTAAGGGCAGAACACC | GTTCACCCAGACATTCGTTAGT |
| 16 | AGGACAAGGAGAGGATTCTCTG | GGAATGGGGGCTTTCAGA |

**Table 6.4 –** Primers used in the amplification and sequencing of *MUTYH*

| Exon | Forward primer sequence | Reverse primer sequence |
|---|---|---|
| 1 | CTTTGGGCGTCGACGAG | GAGGGGACAGGCTTCTCAG |
| 2 (1) | ATTGAGTGCCAGGGTTGTCA | CGGAACCCCAGTGGTGATAC |
| 2 (2) | TGTACTAGCGGATCAAGTAT | TGGCAAAACTGAGTCATAG |
| 3 (1) | GTCTGGTGTTGCTTTCTCTAAC | GTGATGCGGGCGATGTT |
| 3 (2) | TCTCCAGGTGTGCGACTGC | AGGAAGCCTTGAGAAGGTAACC |
| 4 | GGAAGAACTTGAAGATGCCT | GCTCATTTCCTGCTCTCC |
| 5 | CCGGCTTTGGGGCTATA | GTTTCTACCATCCCAGCCCA |
| 6 (1) | TACTTCTGTTGATGGGTCAC | TGGAGGAGAGGAAACCTAG |
| 6 (2) | AGTCACCTCTCCCTCAGACC | GGCTGGAGAGTCCTTTAGGG |
| 7 (1) | ACCTCCCAACACTGTCACTA | TGAACCGGGAGTTTCTCTGC |
| 7 (2) | AGGCTTAGCACTTGCACTTCC | AGGGATCCTTACTGAAGGAC |
| 8 (1) | CTGTGGCCCACGCACTTGTG | ACGTCCTTGGTCCAGCAGTGGT |
| 8 (2) | GAGAGGGGATTCACAAGGTG | GCCATTAGCTCCAGGCTTAC |

**Table 6.5 –** Primers used in the amplification and sequencing of *OGG1*

| Exon | Forward primer sequence | Reverse primer sequence |
|---|---|---|
| 2 | GCAAGGACAGAGGGCTTTCTG | CCAGCAGGCCATCAACTGAT |
| 3 | GCACGTCATGGCTGACTCT | CTGGGAAAGCCGGTTCTAT |
| 4 | TCCCTGGGCTGTGTGTAGAT | GAGATGGGACCCGCATAGT |
| 5 | TGAAGTTTGGGTTGCACCTC | AGATGGTTTGCGGCTGTTC |

**Table 6.6 –** Primers used in the amplification and sequencing of *MTH1* (exon 1 non-coding).

## 6.2.8 Cloning

Three patients carried potential bi-allelic mutations (*MUTYH* G396D and R426L in the first, *MUTYH* G396D and A473T in the second and *OGG1* G308E and A288V in the third). We used cloning to determine if the variants were on the same or opposite DNA strands. Purified amplification products were transformed into high competence JM109 cells which were grown on agar plates. DNA from single colonies was extracted and sequenced using the ABI 3100 system. Sequences were viewed using Sequencher software.

## 6.2.9 *in silico* analysis of variants

We used a combination of computational programs that predict the functional consequence of amino acid changes based on complex algorithms. These were Polyphen (http://genetics.bwh.harvard.edu/pph/), Align GVGD (http://agvgd.iarc.fr/index.php) and SIFT (http://sift.jcvi.org/). Determination of the conservation of a given sequence between various organisms was carried out using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

## 6.2.10 Somatic analyses

Where available, tumour DNA from samples of interest was analysed for somatic mutations of *OGG1*. Tumour DNA was extracted as described in chapter 3.

## 6.2.11 Statistical analysis

Differences between the proportions of cases and controls with each variant were analysed using the Fisher's exact test.

## 6.3 Results

### 6.3.1 Use of the MUTYH variants as controls

Genotyping of the MUTYH variants Y179C and G396D failed on the Illumina golden gate platform. We proceeded to genotype them using a combination of restriction digest and TaqMan assay methods. Three hundred and sixty seven samples were genotyped via both methods to ensure accurate genotyping, showing 100% genotype concordance. The genotype success rates for Y179C and G396D were 99.7% (4351/4362 samples were successfully genotyped) and 99.5% (4340/4362

168

samples successfully genotyped), respectively. In total, only 27 (0.6%) aCRC samples failed genotyping.

### 6.3.2 Analysis of *MUTYH*

Amongst the 2,159 successfully analysed aCRC samples, we identified 11 heterozygotes and one homozygote for Y179C, 33 heterozygotes and one homozygote for G396D, and no compound heterozygotes (Table 6.7). We found similar numbers of Y179C and G396D heterozygotes in cases and controls (no controls were homozygous or compound heterozygous for the *MUTYH* mutations). All heterozygous samples (cases and controls) were sequenced for the entire *MUTYH* open reading frame (ORF) to identify potential biallelic mutations. Apart from the well characterised common polymorphisms V22M and Q338H, we identified a single case of R426L in exon 13 and A473T in exon 14, in different patient samples that also carried G396D.

### 6.3.3 Cloning and sequencing of *MUTYH*

Cloning and sequencing of the genomic regions spanning exons 13 and 14 in the relevant patient samples showed that both R426L and A473T were on opposite strands to G396D indicating that these patients were R426L/G396D and A473T/G396D compound heterozygotes, respectively. Upon review of the pathology notes, the patient homozygous for Y179C and the R426L/G396D compound heterozygote were reported as having 'multiple adenomas' and 'multiple polyps', respectively. No information was available on the two remaining patients. In total therefore, we identified four biallelic *MUTYH* patients (the R426L/G396D and A473T/G396D compound heterozygotes and the Y179C and G396D homozygotes) across our cohort of 2,186 cases.

### 6.3.4 *in silico* analysis

R426L was predicted as likely to have a functional effect by Align-GVGD and to be 'possibly damaging' by Polyphen and A473T was predicted to have a possible effect by Align-GVGD but tolerated by Polyphen.

### 6.3.5 Identifying novel CRC predisposition alleles

Of the 349 SNPs that passed our exclusion criteria, 193 (55.3%) failed *in silico* conversion and 30 (8.6%) were analysed on the chip but the assay failed, leaving 126 (36.1%) SNPs successfully genotyped. The majority of SNPs that failed *in silico* analysis did so either due to their close proximity (<60bp) to other SNPs analysed on the chip or the prediction software deemed them likely to fail analysis. Sixty-one, 70 and 74 SNPs were successfully analysed from the BER, MMR and OxDR pathways respectively.

### 6.3.6 Bi-allelic variants

In total, 13 genes from the three repair pathways were found to contain two or more mutations in cases. Of the 11 involved in BER, 9 (*APEX1, LIG1, MBD4, NEIL1, NEIL3, PARP2, ERCC6, POLG, TSHZ3*) had combinations that were also seen in controls thus excluding these combinations from having a role in CRC predisposition. All 6 of the genes involved in MMR (*EXO1, LIG1, MBD4, MSH4, MSH5, POLε*) had biallelic combinations that were observed in both cases and controls. For the OxDR pathway, 4 of the 6 genes (*APEX1, NEIL1, NEIL3, XRCC1*) had combinations observed in both cases and controls. No combinations of variants were unique to controls.

Two genes were found to contain a combination of two rare nonsynonymous variants that were only found in cases with CRAs or CRCs and never found together in controls. These were *OGG1* and *MTH1*. Details of the biallelic variants, and their biallelic nature, are presented in Table 6.7. These two genes are involved in the BER and the OxDR mechanisms.

| Gene Name | Genotype | aCRC cases (n=2186) | UKBS controls (n=2176) |
|---|---|---|---|
| | Y179C/wt | 11 | 10 |
| | Y179C/Y179C | 1 | 0 |
| MUTYH | G396D/wt | 33 | 29 |
| | G396D/G396D | 1 | 0 |
| | Y179C/G396D | 0 | 0 |
| OGG1 | G308E/wt | 35 | 15 |
| | G308E/G308E | 1 | 0 |
| MTH1 | V83M/wt | 24 | 20 |
| | V83M | 1 | 0 |

Table 6.7 - Genes with combinations of two rare nonsynonymous variants that were only found in CRC cases and never found together in controls. wt – wild-type

### 6.3.7 Biallelic variants of *OGG1*

The *OGG1* variant G308E (rs113561019) was found to exist in a homozygous state in one case but no controls.

### 6.3.8 Analysis of *OGG1*

All G308E heterozygous samples (CRC and CRA cases and controls) were sequenced for the entire *OGG1* ORF and splice regions to identify potential biallelic mutations. In total we identified 4 variants across the ORF of *OGG1*. These were the exon 3 variant R134Q which was observed in one control sample, the exon 5 variant A288V which was observed in one case sample, the common exon 6 variant P332A which was observed in 6 cases and 4 controls, and the synonymous P477P variant which was seen in 4 cases and 1 control.

### 6.3.9 Cloning and sequencing of *OGG1*

Cloning and sequencing of the genomic region spanning exons 5 and 6 in the relevant patient samples showed that A288V was on the opposite strand to G308E, indicating that this patient was a A288V/G308E compound heterozygote. Therefore, we identified two biallelic *OGG1* patients (A288V/G308E and the G308E homozygote) as compared to no biallelic controls.

### 6.3.10 Somatic sequencing of *OGG1*

We sought evidence of a second hit of *OGG1* by screening 23 samples carrying heterozygous G308E (tumour material was not available for the remaining samples). The A400T variant was found in one case. The functional consequence of this variant is unknown.

### 6.3.11 *In silico* analysis

G308E was predicted to be damaging by all *in* silico analysis programs used. Conservation of the base was compared across a series of species including yeast, mouse, bovine, drosophila and humans (Figure 6.1). It was found to be conserved across all species considered. We also carried out *in silico* analysis of the A288V variant and found that it was predicted to be benign and damaging by Polyphen and AlignGVGD respectively. The same analysis of the somatic variant A400T found that it was predicted to be benign and likely to be damaging.

172

### 6.3.12 *OGG1* overrepresentation and validation in the POPGEN population

We compared the frequency of G308E in aCRC cases (1 homozygous and 35 heterozygous) and controls (0 homozygous and 15 heterozygous) and found that it was overrepresented in aCRC ($X^2$=9.6, P=0.002). In addition, we screened for G308E in *OGG1* in 163 patients with CRAs and 480 healthy controls. We identified 3 cases carrying this variant as compared to 1 control sample (P=0.05), suggesting that this variant may also predispose to CRAs.

To clarify the contribution of *OGG1* G308E to CRC risk, we attempted to validate the initial association in a separate cohort. We analysed 2,169 samples from unrelated patients with CRC from the POPGEN cohort. Of the 2,122 successfully analysed cases, 15 were heterozygous and none were homozygous mutant for G308E. Of the 2,741 successfully analysed controls, 26 were heterozygous and none were homozygous mutant. Therefore G308E was not overrepresented in this setting ($X^2$=0.8, P=0.362)

### 6.3.13 Pattern of somatic mutagenesis

We reviewed data from the COIN trial (Chapter 3) to determine if a pattern of somatic mutagenesis was present in the two patients carrying biallelic mutations of *OGG1*. Both patients carried the G12D mutation of *KRAS* which is a G>A transition.

173

```
Human           WHIAQRDYSWHPTTSQ KGPSPQTNKE-------------LGNFFRSLW PYAGWAQAVL 318
Orang-utan      WQIAQRDYSWHPTTSQ KGPSPQTNKE-------------LGNFFRSLW PYAGWAQAVL 318
Rhesus Macaque  WQIAQRDYSWHPTTSQ KGPSPQSNKE-------------LGNFFRSLW PHAGWAQAVL 318
Cow             WQIAQRDYSWHPTTSQ KGPSPQANKE-------------LGNFFRNLW PYAGWAQAVL 320
Dog             WQIAQRDYSWHPSTSQ KGPSPQTNKE-------------LGNFFRSLW PYAGWAQAVL 308
Mouse           WQIAHRDYGWHPKTSQ KGPSPLANKE-------------LGNFFRNLW PYAGWAQAVL 318
Rat             WQIAHRDYGWQPKTSQTKGPSPLANKE-------------LGNFFRNLW PYAGWAQAVL 318
Chicken         WHIARQRYGAALG--- RSLTARVHQE-------------IGDFFRELW PYAGTAE--- 297
Frog            WQVAKRDYLPQLGS-GNKTLTDRVYRE-------------TGDFFHNLW PYAGWAQSVL 357
Zebra-fish      WQIAKRDYNFAPGT-SQKTLTDRVYKE-------------IGDFYRKLW PYAGWAQSVL 344
Salmon          WQIAKRDYSCAAGN-GQKSLTDKVHRQ-------------IGDFFRQLW PYAGWAHSVL 345
Drosophila      YRIAQN--YYLPHLTGQKNVTKKIYEE-------------VSKHFQKLH KYAGWAQAIL 317
Arabidopsis     WQIATN--YLLPDLAG-AKLTPKLHGR-------------VAEAFVSKY EYAGWAQTLL 327
Yeast           SRIAKRDYQISANKNHLKELRTKYNALPISRKKINLELDHIRLMLFKKW SYAGWAQGVL 323
Fungus          YQIAVRDYKFKGNK-SMKTLNKETYAA-------------IRLFFKDIF EYAGWAQSVL 306
                 ::*  .                                           :** *.
```

**Figure 6.1** - Conservation analysis of the *OGG1* A288V (green) and G308E (red) variants across a series of species. An amino acid conserved across all species is indicated with an asterisk (*). Though not shown, conservation analysis of the rest of the gene was carried out and residues conserved across all species tested are relatively uncommon (45/345 human residues).

## 6.3.14 Biallelic variants of *MTH1*

The *MTH1* variant V83M (rs4866) was found to exist in a homozygous state in one case. The SNP was also found in a heterozygous state in 24 cases (CRC) and 20 controls (MAF of 0.5%). This SNP was not overrepresented in CRC cases as compared to controls ($X^2$=0.8, P=0.381) nor in CRAs.

## 6.3.15 Analysis of *MTH1*

We sequenced the ORF and splice regions of V83M heterozygous cases (CRC and CRA) and controls. Besides a synonymous aspartic acid variant at codon 142 (D142D) which was observed in 3 cases and 5 controls, no new variants were observed.

## 6.3.16 *In Silico* analysis

We carried out *in silico* analysis of the *MTH1* variant V83M. It was found to be conserved across human, bovine and mouse. V83M was predicted to be benign by both Polyphen and AlignGVGD.

## 6.4 Discussion

### 6.4.1 *MUTYH* proof of principle

The established pathogenic *MUTYH* variants Y179C and G396D were included in this analysis as 'proof-of-principle' controls. We found two CRC cases that were homozygous for these variants and proceeded to identify two further patients that were R426L/G396D and A473T/G396D compound heterozygotes. No controls contained a biallelic combination of *MUTYH* variants. Whilst the number of biallelic mutation carriers was less than expected, we nonetheless prove that this screening approach would have highlighted the importance of rare *MUTYH* variants in predisposition to CRC.

### 6.4.2 Prognostic effect of MAP

We compared the frequency of patients with biallelic *MUTYH* mutations in the aCRC setting to previous studies of CRC patients with earlier stages of disease. Cleary and colleagues (2009) studied 3,811 cases with CRC collected from a multisite CRC registry from Canada, the United States and Australia. In total, 42.8% of their cases

had Stage 1 disease, 30.1% had Stage 2 disease, 22.2% had Stage 3 disease and 5.6% had Stage 4 disease/aCRC. Overall, they identified a total of 27 cases with biallelic *MUTYH* mutations, which included one of the 212 patients with aCRC. We found that patients with biallelic *MUTYH* mutations were much rarer in the advanced as compared to the earlier (Stages 1-3) disease setting (4/2,159 vs. 26/3,623, $P$=3.58x10$^{-3}$, using our aCRC data; and 5/2,371 vs. 26/3,623, $P$=4.53x10$^{-3}$ with the aCRC data combined).

Other large population based studies of CRC have also screened for *MUTYH* mutations and although none give comprehensive breakdowns of their disease spectrum, it is likely that both have distributions similar to the registers described above. Seventeen patients with biallelic *MUTYH* mutations were identified amongst 3,145 Scottish CRC cases (Farrington *et al.* 2005; Tenesa *et al.* 2006) ($P$=3.18x10$^{-2}$ compared to the advanced disease setting) and 8 patients with biallelic *MUTYH* mutations were identified amongst 1,116 Spanish CRC cases ($P$=2.14x10$^{-2}$ compared to the advanced disease setting). Although Lubbe and colleagues (2009) identified 27 patients with biallelic *MUTYH* mutations amongst 9,268 British CRC cases (which is not significantly different compared to the advanced disease setting), this study did not screen their one-hundred and ninety-eight Y179C or G396D single heterozygotes for biallelic mutations, so the true number of biallelic *MUTYH* cases is likely to be significantly under-represented. Therefore, patients with biallelic germline *MUTYH* mutations have a significantly improved prognosis as compared to patients that do not carry these mutations.

### 6.4.3 *OGG1*

The *OGG1* gene was found to carry biallelic variants in two aCRC cases and in no controls. One case was a carrier of homozygous G308E, while the other was a compound heterozygote for G308E and A288V. OGG1 plays a critical role in the BER process, excising the 8oxoG lesion from the 8oxoG:C duplex that results from oxidative damage. Given this role in a repair pathway that has already been linked with CRC predisposition, *OGG1* represented an excellent candidate for further study.

G308E was first observed in a screen of *OGG1* in patients with head and neck cancer (Blons *et al.* 1999). Glycine 308 is a conserved amino acid in a range of

176

diverged species (human to yeast to drosophila). Blons and colleagues produced GST fusion proteins (representing the normal and variant protein) and tested their activity through *in vitro* and *in vivo* substrate cleavage assays. They found that purified protein containing the variant displayed an ability to cleave a double-stranded DNA 34-mer containing an 8oxoG:C base pair, that was equivalent to protein containing the wild-type amino acid sequence. Beyond this initial data, little is known about the effect of G308E though it results in the substitution of an aliphatic residue with an acidic residue and is predicted to be damaging by a range of prediction software.

A288V has been linked with reduced glycosylase activity. In their study of Alzheimer's patients, Mao and colleagues (2007) screened 14 patients for *OGG1* mutations. One of the four variants they identified was A288V which showed decreased glycosylase activity (50-60% activity as compared to wild-type protein). They suggest that this variant probably functions by reducing the affinity of the glycosylase for the DNA substrate. Indeed the crystal structure of the protein would suggest this as alanine 288 comes into direct contact with DNA. A recent kinetic analysis of this variant (Sidorenko *et al.* 2009) confirmed reduced protein activity (~30% less efficient compared to wild-type) through an oligonucleotide cleavage assay. This was not, however, confirmed by an irradiated DNA assay where its cleavage capacity was comparable to wild-type protein. The 3D locations of this variant, and G308E, are highlighted in Figure 6.2.

We analysed COIN trial data (chapter 3) to determine if a pattern of somatic mutagenesis was observed in the two patients carrying a biallelic combination of variants. Based on the functional overlap OGG1 has with MUTYH, we expected to see G to T transversions. Instead we observed that both cases carried G to A transitions in *KRAS*. The significance of this is unknown though G to A transitions can result from 8oxoG:T mispairs which are a secondary target of OGG1 (Bjørås *et al.* 1997). Bjørås and colleagues suggest that there are potential issues for strand discrimination when OGG1 encounters less common mispairs such as 8oxoG:T and it would be interesting to see if G308E is capable of influencing this.

OGG1 G308E was initially found to be over-represented in aCRC cases as compared to healthy controls ($X^2$=9.6, P=0.002). Despite this, we did not show an association in an independent cohort of 2,713 histologically proven colorectal carcinoma patients and 2,145 controls ($X^2$=0.8, P=0.366). It is possible that the effects of OGG1 G308E are specific to the metastatic/advanced diseased state and as such this requires further investigation.

### 6.4.4 *MTH1*

The *MTH1* V83M variant was found to exist in a biallelic state (i.e. homozygous) in one case and no controls. Like OGG1, MTH1 has a critical role within the BER process, functioning to remove 8-oxodGTP from the nucleotide pool. Interestingly, V83M has been associated with an increased risk of small cell lung carcinoma (Kohno *et al.* 2006) and gastric cancer (Kimura *et al.* 2004). V83M containing protein has been reported to be structurally and catalytically more thermolabile than the wild-type protein (Yakushiji *et al.* 1997). Coincident with its association with gastric cancer, the presence of the 83M variant correlates with somatic mutation of the *p53* tumour suppressor gene, specifically G:C→T:A and A:T→C:G transversions (Tsuzuki *et al.* 2001b; Kimura *et al.* 2004). Interestingly, our patient that was homozygous for V83M was found to have a somatic G>T mutation in *KRAS*. The role of this variant in CRC predisposition requires further study.

### 6.4.5 Future work

Based on the strong biological rational that already exists for a role of BER pathway components in CRC risk, it will be important to determine if the biallelic mutations observed in OGG1 and MTH1 do alter protein function and in-turn risk. This will require the use of a series of functional experiments to determine if these mutations alter expression, repair capacity, localisation etc. Of note was the fact that OGG1 G308E was over-represented in cases vs. controls within the advanced disease setting. An important next step will be to attempt to validate this observation in an independent cohort of advanced cases.
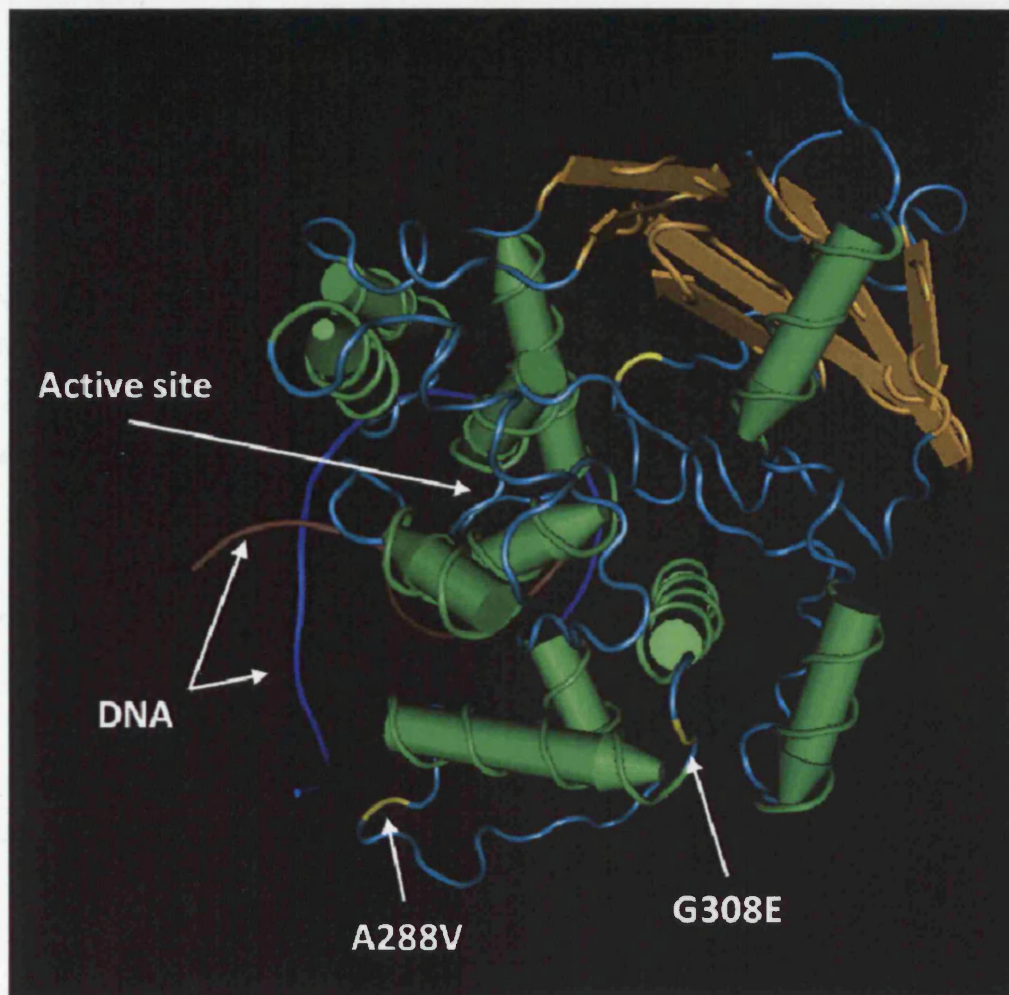
**Figure 6.2** - Crystal structure of the OGG1 protein cross-linked with guanine-containing DNA. The A288 and G308 amino acids are highlighted in yellow and indicated by arrows. A288 comes into direct contact with DNA (brown and blue strands) with the A288V variant influencing catalytic activity. G308E lies on the surface of the protein and has no obvious functional consequence.

# Chapter Seven - General Discussion

## 7.1 Factors affecting patient survival (Table 7.1)

### 7.1.1 Germline factors

The data presented in this thesis represent the first report of a role for germline CRC susceptibility alleles in influencing patient survival. In all, prognostic roles have been identified for four of the fourteen currently known GWAS loci. These were validated in an independent cohort of patients. The four loci are 16q22 (rs9929218), 19q13 (rs10411210), 10p14 (rs10795668) and 8q23 (rs16892766). Interestingly, others (Tenesa *et al.* 2010; Cicek *et al.* 2009; Xing *et al.* 2011) have scrutinised GWAS risk loci for prognostic roles but have reported no convincing evidence for prognostic influence. It is likely that the discordance between our findings and those of Tenesa and colleagues lies in the difference in stage of patients screened, with the latter using patients with largely stages 1-3 CRC. As such, they likely had less power to observe effects over OS. Of note, however, the prognostic effects of rs16892766, rs9929218 and rs10411210 observed here were not significant across all trial arms. Since different amounts of the chemotherapeutic backbone were used across these arms, we cannot rule out that the difference in survival outcome is a result of altered response to therapy.

It is likely that the upcoming years will see GWAS technology employed in the identification of genetic factors associated with CRC survival. GWAS for survival alleles have been carried out in other cancers, specifically breast (Azzato *et al.* 2010) and prostate (Penney *et al.* 2010). Neither study managed to identify germline variants though both had only limited power to detect modest effects on survival for common variants.

### 7.1.2 Mechanism at the 8q23 locus

rs16892766 lies downstream of the *EIF3H* gene, at 8q23, which has previously been shown to have oncogenic potential (Zhang *et al.* 2007). Further scrutiny of this gene revealed another SNP (rs28649280) that was in moderate LD with rs16892766 and that also acted as a susceptibility allele. Furthermore, *in silico* analysis suggested that rs28649280 lies in the promoter region of the *EIF3H* gene, in a binding site for the SP1 transcription factor. Using a luciferase assay we found that there was an

allele specific effect upon expression – with the risk allele associating with significantly increased luciferase activity. Of note, both rs28649280 and the rs16892766 carried a negative prognostic effect on OS. It is possible that a multitude of elements individually and collectively regulate EIF3H expression, including the rs28649280 and rs16888589 SNPs described by ourselves and Pittman and colleagues (2010) (Figure 7.1). When one or more risk alleles are present, they modestly alter expression of EIF3H which subtly contributes to the progression of CRC through increased expression of transcription factors and proto-oncogenes. Ultimately, this alters the risk and prognosis of the disease.

### 7.1.3 Distinct mechanisms of colorectal tumourigenesis and their prognostic influence

We have shown a scarcity of biallelic carriers of *MUTYH* variants in the advanced disease setting. We suggest that this rarity may be explained by a significantly improved prognosis for these patients as compared to those without *MUTYH* mutations. Improved survival had previously been shown in a comparison of 147 MAP patients with CRC and 272 matched patients with CRCs (five year survival being 78% and 63% respectively) (Nielsen *et al.* 2010). This observation was significant (HR 0.48, 95% CI 0.32-0.72, P<0.001) after adjustment for multiple factors including age, stage, sex, tumour sub-site, country and year of diagnosis. Closer scrutiny of prognostic impact for stage specific cancers revealed that the survival benefit for MAP patients compared with controls was greater for patients with early-stage CRC (HR 0.45, 95% CI 0.23-0.91) than those with later stage disease (HR 0.64, 95% CI 0.34-1.20). The authors acknowledge that there are a number of weaknesses to this study, including multiple selection biases (including the earlier clinical attention given to members of 'predisposed' families, polyp status of MAP cases and discrepancies between time of diagnosis for MAP and population based CRC sufferers) and differences in immune response. Nonetheless, the data points to a significantly better survival for MAP patients than for matched control patients with CRC. The MAP data presented in this thesis combined with that of Nielsen and colleagues (2010) and data from HNPCC-associated tumours and sporadic CRCs with MSI indicate that distinct mechanisms of colorectal tumourigenesis may directly influence patient survival.
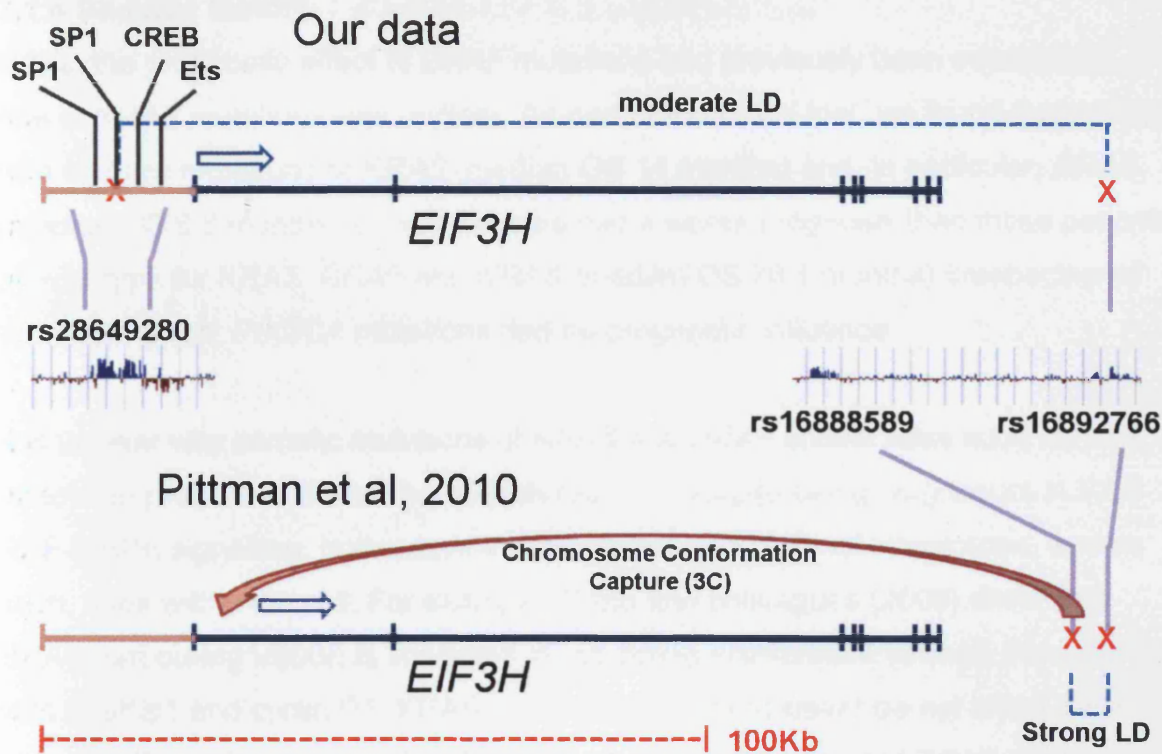
**Figure 7.1** - Diagrammatic representation of two potential EIF3H regulatory mechanisms. The main transcript of *EIF3H*, and its exons, is represented with a blue line drawn to scale. SNPs are shown with a red cross. The EIF3H promoter region is represented by a light purple line that is not drawn to scale. Conservation at various sites is shown by the blue and red jagged line – a blue peak highlights areas that are heavily conserved. We propose a model in which rs16892766 is in moderate LD with another SNP that physically lies in the promoter region, rs28649280. This SNP lies in a conserved potential binding site for the SP1 transcription factor and alters expression. In the mechanisms proposed by Pittman and colleagues (2010), a SNP (rs16888589) in high LD with rs16892766 was found to lie in a conserved enhancer element. Furthermore, they show that this conserved island physically interacts with the promoter region of the EIF3H gene.

## 7.1.4 Somatic factors

Whilst the prognostic effect of *BRAF* mutations had previously been established, the role of *KRAS* mutations was unclear. As part of the COIN trial, we found that patients with somatic mutations of *KRAS* (median OS 14 months) and, in particular, *BRAF* (median OS 8.8 months) in their tumours had a worse prognosis than those patients all wild-type for *KRAS, BRAF* and *NRAS* (median OS 20.1 months) irrespective of treatment given. *PIK3CA* mutations had no prognostic influence.

It is unclear why somatic mutations of *KRAS* and *BRAF* should have such varying effects on prognosis. It must be considered that, despite being neighbours in RAS-RAF-MAPK signalling, both enzymes have differing functional interactions, and as such, roles within the cell. For example, Preto and colleagues (2008) show that BRAF harbouring V600E is important in regulating proliferation through interactions with p29Kip1 and cyclin D1. KRAS harbouring G13D however do not affect these proteins. They also suggest that CRC tumour cells with activated BRAF are more dependent on the BRAF-ERK pathway for proliferation than those with activated KRAS which may signal to other pathways, including the PI3K-PTEN-Akt and RalGDS pathways (Solit *et al.* 2006; Schubbert *et al.* 2007).

It is likely that the particular interactions and mechanisms that selectively occur in *BRAF* mutant tumours are conducive to a more aggressive phenotype than in tumours harbouring mutations of *KRAS*. This ultimately reduces the survival time of patients with this somatic mutation. Similarly, the course taken by tumour with a *KRAS* mutation is likely to be more aggressive than that taken by a tumour without a *KRAS* mutation. The complexity of prognostic roles for the somatic phenotypes is further highlighted by the fact that the negative prognostic value of mutant *BRAF* may be restricted to patients that are MSS (Roth *et al.* 2010). Therefore the molecular signature of a tumour influences patient survival – with different mutations affecting individual and interacting pathways.

## Factors affecting patient survival

| Loci | SNP/Gene | Mechanism | HR | 95% CI | P value | Survival Difference (median months) |
|------|----------|-----------|-----|--------|---------|-------------------------------------|
| 8q23 | rs16892766 | Altered *EIF3H* expression | 1.27 | 1.12-1.44 | <0.001 | -2.9 |
| 8q23 | rs28649280 | Altered *EIF3H* expression | 1.26 | 1.12-1.43 | <0.001 | -2.9 |
| 8q23 | rs16888589 | Altered *EIF3H* expression | 1.21 | 1.07-1.37 | 0.002 | -2.7 |
| 16q22 | rs9929218 | Altered *CDH1* expression | 1.46 | 1.23-1.74 | 0.002 | -4.3 |
| 16q22 | rs16260 | Altered *CDH1* expression | 1.38 | 1.15-1.67 | 0.001 | -3.9 |
| 19q13 | rs10411210 | Unknown | 1.23 | 1.08-1.39 | 0.001 | -2.5 |
| 10p14 | rs10795668 | Unknown | 0.71 | 0.59-0.86 | 0.001 | +6.2 |
| 12p12 | *KRAS* | Constitutive Ras-Raf-MAPK signalling | 1.34 | 1.18-1.52 | <0.001 | -3.2 |
| 7q34 | *BRAF* | Constitutive Ras-Raf-MAPK signalling | 2.00 | 1.61-2.50 | <0.001 | -7.9 |
| 1p34 | *MUTYH** | Impaired ability to repair G:C→T:A transversions. | - | - | - | - |

**Table 7.1** – Summary of factors affecting patient survival studied in this thesis. *patients with biallelic *MUTYH* mutations associate with a significantly improved prognosis as compared to patients without these mutations.

## 7.2 Factors affecting response to and side effects from chemotherapy (Table 7.2)

We tested all SNPs across the 14 CRC risk loci for their effects on response to, and side effects from 12 weeks of standard oxaliplatin and fluoropyrimidine based chemotherapy. We found that patients homozygous for the rs9929218 minor allele had a poor prognosis for response to treatment as compared to patients' heterozygous or homozygous wild-type. These patients also had an increased risk of skin toxicity, specifically HFS. The strongly linked variant rs16260 had a similar effect on response and side effects. Though rs16892766 did not have an effect, we found that rs28649280 and rs16888589 at 8q23 did alter response to treatment at 12 weeks.

It is unclear as to why these variants should alter response to therapy. There is no data to suggest why EIF3H might alter response and no obvious biological rationale. However, it has been suggested that the levels of E-cadherin expression can alter response to cytotoxic agents (including oxaliplatin and 5FU) in breast cancer, with cells expressing E-cadherin showing a trend towards improved response (Koo *et al.* 2009). Meanwhile, studies of CRC cells that developed resistance to oxaliplatin showed that they underwent EMT, with an associated reduction of E-cadherin expression (Yang *et al.* 2006).

**Factors affecting response to and side effects from chemotherapy**

| Loci | SNP/Gene | Parameter | Mechanism | OR | 95% CI | P value |
|------|----------|-----------|-----------|-----|--------|---------|
| 8q23 | rs28649280 | Response | Altered *EIF3H* expression | 0.60 | 0.43-0.83 | 0.002 |
| 8q23 | rs16888589 | Response | Altered *EIF3H* expression | 0.72 | 0.54-0.97 | 0.031 |
| 16q22 | rs9929218 | Response | Altered *CDH1* expression | 0.48 | 0.31-0.75 | 0.001 |
| | | HFS | | 4.68 | 1.84-11.90 | 0.001 |
| 16q22 | rs16260 | Response | Altered *CDH1* expression | 0.58 | 0.37-0.93 | 0.022 |
| | | HFS | | 4.78 | 1.80-12.7 | 0.002 |

**Table 7.2** – Summary of factors affecting response to and side effects from chemotherapy studied in this thesis.

## 7.3 Predictive factors affecting response to cetuximab (Table 7.3)

### 7.3.1 The COIN trial

The negative outcome of the COIN trial highlights how much there is yet to learn about the correct use of cetuximab in the management of mCRC. Against expectations, the trial showed no benefit for cetuximab administration in patients that were wild-type for *KRAS*. This was also the case in patients all wild-type for *KRAS*, *BRAF* and *NRAS*. The reasons for this negative outcome remain unclear. It is possible that the apparent lack of benefit from cetuximab is merely an artefact of reduced fluoropyrimidine application (the dose of capecitabine was reduced due to toxicity issues). Also the COIN data suggests that capecitabine, a popular choice of fluoropyrimidine since it is administered orally, may interact with cetuximab. This is possible since the drugs have overlapping toxicity profiles; Cetuximab induces increased diarrhoea, potentially impacting on the amount of capecitabine that is absorbed through the gut. A recent comparison of cetuximab plus OxCap and cetuximab plus CAPIRI (capecitabine + irinotecan) in 177 mCRC patients found no improvement in treatment efficacy in *KRAS* wild-type patients as compared to *KRAS* mutant (Moosmann *et al.* 2011) in the OxCap group. Though this was an under-powered study with no control arm, this data compliments the COIN trial results. However, it should be noted that the authors did not observe increased nonhematological or gastrointestinal toxicity with OxCap suggesting the interaction between agents may be a mechanistic one.

If patients that received capecitabine are excluded, the COIN trial data more closely resembles that of the majority of other clinical trials of cetuximab efficacy. A most responsive cohort was identified that contained patients wild-type for *KRAS*, *BRAF* and *NRAS*, contained ≤1 metastatic sites and received OxFU and these patients did show improved PFS in response to cetuximab treatment (HR 0.55, 95% CI 0.35-0.87, P=0.01).

Though data has not yet been published, the results from the NORDIC VII study have recently been presented at the ESMO 2010 meeting. This trial tested the addition of cetuximab to NORDIC FLOX (5FU/Folinic Acid and Oxaliplatin) in mCRC patients. NORDIC VII, like COIN, showed no statistically significant difference in PFS, OS, or ORR between patients that received cetuximab and those that did not

even when considering *KRAS* status. Of note, no patients were administered capecitabine suggesting that oxaliplatin, and not capecitabine, may be contributing a lack of response to cetuximab.

## 7.3.2 The role of E-cadherin in response to cetuximab

Our study of risk loci for predictive influences revealed that the rs9929218 polymorphism at 16q22 confers altered response to cetuximab and risk of HFS at 12 weeks. This variant maps to the second intron of the *CDH1* gene which encodes E-cadherin. E-cadherin is a transmembrane glycoprotein that is critical in the establishment and maintenance of intercellular adhesion, cell polarity and tissue morphology and regeneration (Takeichi, 1991). It localises at adherens junctions where it interacts, through its cytoplasmic component, with β-catenin. This in turn associates with α-catenin which mediates its linkage with the cell's cytoskeletal structure (Tsanou *et al.* 2008). Loss of E-cadherin represents the defining feature of the 'epithelial to mesenchymal transition' (EMT) – a change in cell phenotype that sees cells down-regulate cell-cell adhesion structures, as well as alter their polarity and cytoskeletal structure (Klymkowsky and Savagner, 2009). EMT is a common event during epithelial-derived tumour progression.

rs9929218 lies in a region of CDH1 that contains several cis-regulatory elements (Stemmler et al. 2005) and is in strong LD with rs16260 in its promoter, the minor allele of which down-regulates E-cadherin expression (Li et al. 2000). Patients homozygous for the rs9929218 minor allele (~8% of patients), showed a poor prognosis with standard oxaliplatin and 5FU-based chemotherapy, but showed significant benefit from treatment with cetuximab. Interestingly, these patients would be predicted to have reduced E-cadherin expression (from linkage to rs16260) and consequently expected to have poor response to cetuximab (see chapter 3). Furthermore, we show that the benefit was greatest in those patients that also had *KRAS, BRAF* or *NRAS* mutant tumours, again in contrast with expectations (De Roock et al. 2010a). We therefore suggest that cetuximab may be benefiting these patients by a novel mechanism similar to the synthetic sickness/lethalilty phenotype observed in BRCA1 or BRCA2 mutant cells with PARP1 inhibition, MSH2 mutant cells with POLB inhibition and MLH1 mutant cells with POLG inhibition (Martin et al. 2010) (Figure 7.2A). We propose that a combination of reduced E-cadherin

expression, activation of the Ras-Raf-MAPK signalling pathway and EGFR blockade (via cetuximab) creates a cellular environment incompatible with survival which causes the tumour to shrink (Figure 7.2B). Further studies are therefore warranted to exploit this potential therapeutic opportunity.

## 7.4 Aims of this thesis

In section 1.7, a series of aims were described for this thesis. These aims have been addressed as follows;

**Chapter 03** – Through the COIN trial, we have contributed to the understanding of the role played by somatic mutations within key components of the EGFR signalling pathways, in altering response to cetuximab. Though overall the data suggests that these variants play no role in altering response, it is likely that a number of confounding factors are masking the true effects these mutations play. This analysis allowed the comparison of pyrosequencing and Sequenom for the high-throughput analysis of somatic variants, with results suggesting high genotyping success rate and concordance between both technologies.

**Chapter 04** – The genotyping of COIN patients and UKBS controls allowed for the determination of the role played by a series of variants in defining CRC risk within the advanced disease setting. Through a series of functional experiments, we show that EIF3H may be involved in altering this risk (and patient survival) through its linkage to a SNP in the EIF3H promoter region.

**Chapter 05** – We have identified a series of prognostic effects for SNPs at the 8q23, 16q22, 19q13 and 10p14 loci. Furthermore, exploiting the design of the COIN trial, we show that the rs9929218 SNP alters both prognostic response to standard chemotherapy and predictive response to cetuximab. These effects are also shown by the rs16260 SNP in the CDH1 promoter region.

**Chapter 06** – We aimed to identify novel high penetrance alleles that predispose to CRC from the BER, MMR and OxDR pathways. Genes found to carry biallelic mutation in cases and not controls included MUTYH, OGG1 and MTH1 which operate as a triumvirate of repair genes in the BER pathway. Further work is needed to determine if the OGG1 G308E and MTH1 V83M mutations do contribute to CRC risk. By virtue of its significant under-representation in the advanced disease setting, we suggest that MAP confers a positive prognosis upon CRC survival.
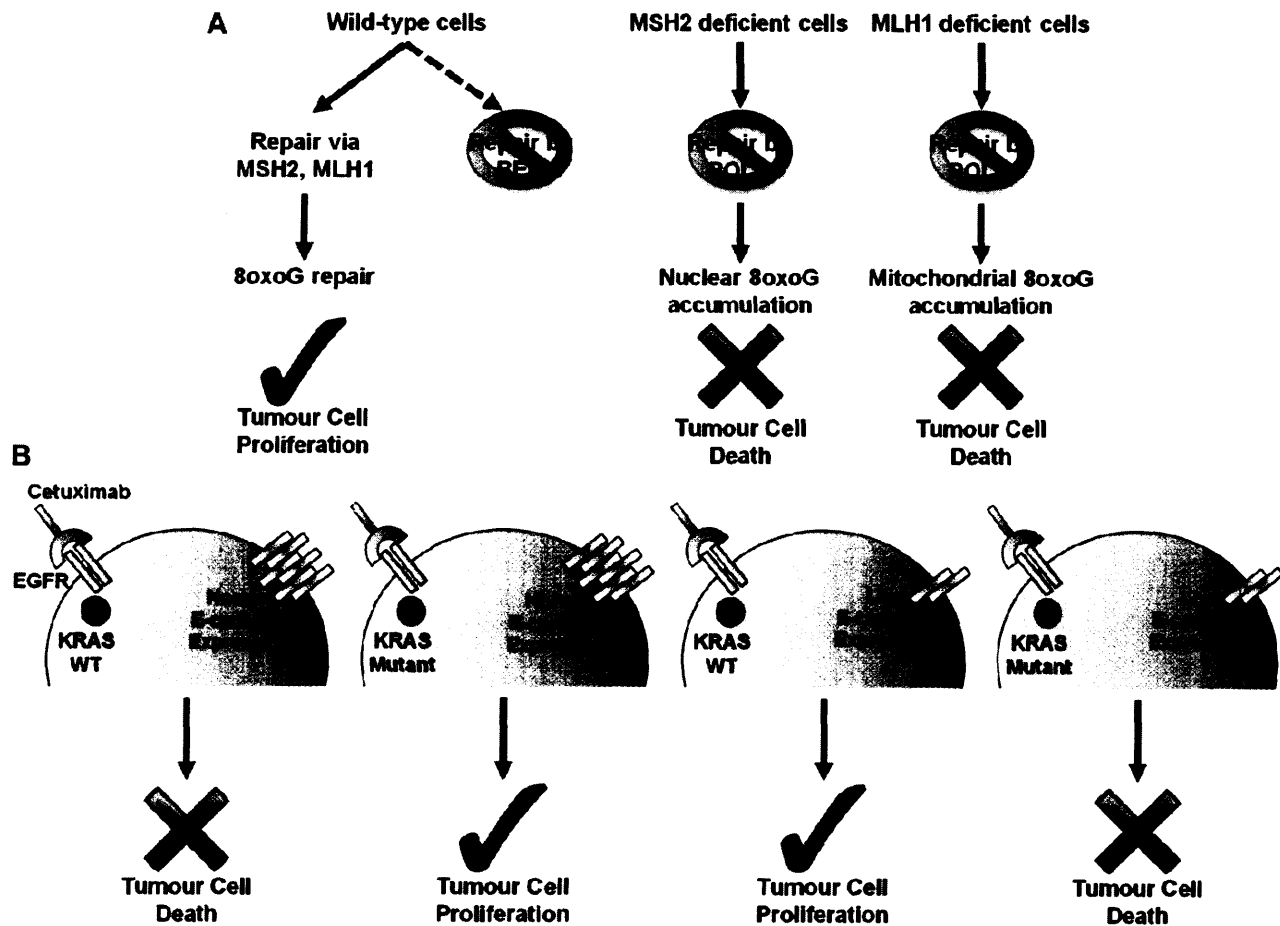
**Figure 7.2** – Synthetic lethality. A, the Mechanism put forward by Martin and colleagues (2010). Here, repair of oxidized DNA lesions can be performed by either MSH2/MLH1 dependent processes or BER. In wild-type cells, inhibition of BER (through POLB and/or POLG) does not result in cell death since repair can be carried out by MSH2/MLH1. In MSH2 mutant cells, inhibition of POLB results in the accumulation of 8oxoG within nuclear DNA which may cause cellular arrest or death. In MLH1 mutant cells, inhibition of POLG results in the accumulation of 8oxoG in mitochondrial DNA which again results in reduced replicative potential or death. B, Diagram representing the possible mechanism through which tumour cells respond to cetuximab in an rs9929218 and *KRAS, BRAF, NRAS* mutant dependent manner. In the first scenario, blockade of EGFR signalling by cetuximab results in death of the tumour cell. However, when *KRAS* is mutated, cetuximab is rendered ineffective and tumour cells proliferate unchecked. In the third scenario, and as suggested by Black and colleagues (2008), cells lacking E-cadherin are resistant to cetuximab. Finally, the combination of reduced E-cadherin expression, activation of the Ras/Raf signalling pathway and EGFR blockade creates a cellular environment incompatible with tumour cell survival.

## Factors affecting patient response to cetuximab

| Aim | Effect | HR/OR | 95% CI | P value | Mechanism |
|---|---|---|---|---|---|
| To determine if there was improved response to cetuximab in CRC patients... | | | | | |
| ...wild-type for KRAS | No effect (OS) | 1.04 | 0.87-1.23 | 0.67 | |
| ...wild-type for KRAS, BRAF, NRAS | No effect (OS) | 1.02 | 0.83-1.24 | 0.86 | There may have been an interaction between capecitabine and cetuximab (perhaps through overlapping toxicities). |
| ...wild-type for KRAS, BRAF NRAS, with ≤1 metastatic site and that received OxFU therapy | Improved response (PFS) | 0.55 | 0.35-0.87 | 0.01 | |
| ...wild-type for PIK3CA | No effect (OS) | 1.01 | 0.88-1.16 | 0.89 | The role of PIK3CA mutations in altering response to cetuximab is unclear. Our data suggests that it does not have an effect, regardless of KRAS status. |
| ...wild-type for PIK3CA and KRAS | No effect (OS) | 1.03 | 0.86-1.24 | 0.75 | |
| Of the 14 CRC risk loci, an effect on response was observed in CRC patients... | | | | | |
| ...homozygous for the minor allele of rs9929218... | Improved response | 1.69 | 0.93-3.07 | 0.001* | rs9929218 lies in 2nd intron of CDH1. It is possible that tumour cells are killed by a synthetic lethality mechanism. Alone mutations of E-cadherin and Ras-Raf signalling are not sufficient to induce tumour cell death, but combined these mutations are incompatible with tumour survival upon EGFR blockade. |
| ...combined with mutant KRAS, BRAF or NRAS | Improved response | 1.69 | 0.61-4.74 | 0.004* | |

**Table 7.3** – Summary of factors affecting response to cetuximab studied in this thesis. * P for interaction.

## Publications resulting from this work

**-Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33**

Richard S Houlston, Jeremy Cheadle, Sara E Dobbins, Albert Tenesa, Angela M Jones, Kimberley Howarth, Sarah L Spain, Peter Broderick, Enric Domingo, Susan Farrington, James G D Prendergast, Alan M Pittman, Evi Theodoratou, **Christopher G Smith**, Bianca Olver, Axel Walther, Rebecca A Barnetson, Michael Churchman, Emma E M Jaeger, Steven Penegar, Ella Barclay, Lynn Martin, Maggie Gorman, Rachel Mager, Elaine Johnstone, Rachel Midgley, Iina Niittymäki, Sari Tuupanen, James Colley, Shelley Idziaszczyk, The COGENT Consortium, Huw J W Thomas, Anneke M Lucassen, D Gareth R Evans, Eamonn R Maher, The CORGI Consortium, The COIN Collaborative Group, The COINB Collaborative Group, Timothy Maughan, Antigone Dimas, Emmanouil Dermitzakis, Jean-Baptiste Cazier, Lauri A Aaltonen, Paul Pharoah, David J Kerr, Luis G Carvajal-Carmona, Harry Campbell, Malcolm G Dunlop & Ian P M Tomlinson

**-Analysis of the frequency of *GNAS* codon 201 mutations in advanced colorectal cancer**

S. Idziaszczyk, C.H. Wilson, **C.G. Smith**, D.J. Adams & J.P. Cheadle

**- The addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial**

Maughan TS, Adams RA, **Smith CG**, Meade AM, Seymour M, Wilson R, Idziaszczyk S, Harris R, Fisher D, Kenny SL, Kay E, Mitchell JK, Madi A, Jasani B, James MD, Bridgewater J, Kennedy MJ, Claes B, Lambrechts D, Kaplan R (corresponding author), Cheadle JP, on behalf of the MRC COIN Trial Investigators

**-Multiple colorectal cancer susceptibility alleles affect patient survival and**
***CDH1* variants predict response to cetuximab**
**Christopher G. Smith**, David Fisher, Rebecca Harris, Timothy S. Maughan, Angela
Meade, Shelley Idziaszczyk, James Colley, Hannah West, COIN and COIN-B
Collaborative Groups, Richard Kaplan and Jeremy P. Cheadle
JNCI *(in submission)*


## Conference proceedings and poster presentations

**-BSHG (Warwick, September 2010) - A novel colorectal cancer susceptibility**
**SNP in the *EIF3H* promoter influences patient survival and response to**
**treatment.**
**Christopher G. Smith**, David Fisher, Rebecca Harris, Timothy S. Maughan, Shelley
Idziaszczyk, Valentina Moskvina, James Colley, Angela Meade, Cleo Bonnet,
Hannah West, Richard Kaplan, The COIN Collaborative Group, The COIN-B
Collaborative Group and Jeremy P. Cheadle

**NCRI (Liverpool, November 2010) - A novel colorectal cancer susceptibility**
**SNP in the EIF3H promoter influences patient survival and response to**
**treatment.**
**Christopher G. Smith**, David Fisher, Rebecca Harris, Timothy S. Maughan, Shelley
Idziaszczyk, Valentina Moskvina, James Colley, Angela Meade, Cleo Bonnet,
Hannah West, Richard Kaplan, The COIN Collaborative Group, The COIN-B
Collaborative Group and Jeremy P. Cheadle

**ESMO (Milan, October 2010) - A novel colorectal cancer susceptibility SNP in**
**the *EIF3H* promoter influences patient survival and response to treatment.**
**Christopher G. Smith**, David Fisher, Timothy S. Maughan, Valentina Moskvina,
Angela Meade, Hannah West, Richard Kaplan, The COIN Collaborative Group, The
COIN-B Collaborative Group and Jeremy P. Cheadle

**ESMO (Milan, October 2010) - High throughput somatic profiling of the Ras-Raf-MAP and PI3K-PTEN-Akt pathways in advanced colorectal cancer and correlations with response to cetuximab**

**Christopher G. Smith**, Bart Claes, David Fisher, Shelley Idziaszczyk, Gilian Peuteman, Rebecca Harris, Natacha Lays, Michelle James, Bharat Jasani, Richard A. Adams, Richard Kaplan, The COIN collaborative group, Angela Meade, Diether Lambrechts, Timothy S Maughan and Jeremy P Cheadle

# References

Aaltomaa, S. *et al.* 1991. Hormone receptors as prognostic factors in female breast cancer. *Ann Med* 23, pp. 643 - 648.

Aaltonen, L. *et al.* 2007. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin Cancer Res* 13: 356–361.

Aburatani, H. *et al.* 1997. Cloning and Characterization of Mammalian 8-Hydroxyguanine-specific DNA Glycosylase/Apurinic, Apyrimidinic Lyase, a Functional mutM Homologue. *Cancer Res*, 57(11), pp. 2151-2156.

Adams, R. *et al.* 2011. Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy in the first-line treatment of patients with advanced colorectal cancer: results of the MRC COIN trial. *Lancet Oncol,* in press.

Alarmo, E.-L. and Kallioniemi, A. 2010. Bone morphogenetic proteins in breast cancer: dual role in tumourigenesis? Endocr *Relat Cancer* 17(2), pp. R123-139.

Albertini, R. J. *et al.* 1990. In Vivo Somatic Mutations in Humans: Measurement and Analysis. *Annual Review of Genetics* 24(1), pp. 305-326.

Albertson, T. M. *et al.* 2009. DNA polymerase ε and δ proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc Natl Acad Sci USA,* 106(40), pp. 17101-17104.

Allegra, C. J. *et al.* 2009. American Society of Clinical Oncology Provisional Clinical Opinion: Testing for KRAS Gene Mutations in Patients With Metastatic Colorectal Carcinoma to Predict Response to Anti–Epidermal Growth Factor Receptor Monoclonal Antibody Therapy. *Journal of Clinical Oncology* 27(12), pp. 2091-2096.-

Allinson, S. L. *et al.* 2003. Poly(ADP-ribose) polymerase in base excision repair: always engaged, but not essential for DNA damage processing. *Acta Biochim Pol*, 50(1), pp. 169-179.

Allred, D. *et al.* 1990. Immunocytochemical analysis of estrogen receptors in human breast carcinomas. Evaluation of 130 cases and review of the literature regarding concordance with biochemical assay and clinical relevance. *Arch Surg* 125, pp. 107 - 113.

Al-Tassan, N. *et al.* 2002. Inherited variants of MYH associated with somatic G:C[rarr]T:A mutations in colorectal tumors. *Nat Genet* 30(2), pp. 227-232.

Amado, R. *et al.* 2008. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* 26(10), pp. 1626 - 1634.

Amador, M. L. *et al.* 2004. An Epidermal Growth Factor Receptor Intron 1 Polymorphism Mediates Response to Epidermal Growth Factor Receptor Inhibitors. *Cancer Research* 64(24), pp. 9139-9143.

Ames, B. N. and Gold, L. S. 1991. Endogenous mutagens and the causes of aging and cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 250(1-2), pp. 3-16.

An, Q. *et al.* 2005. C → T mutagenesis and γ-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *EMBO J*, 24(12), pp. 2205-2213.

Anderson, W. *et al.* 2001. Tumor variants by hormone receptor expression in white patients with node-negative breast cancer from the surveillance, epidemiology, and end results database. *J Clin Oncol* 19, pp. 18 - 27.

Andreyev, H. *et al.* 2001. Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study. *Br J Cancer* 85(5), pp. 692 - 696.

Andreyev, H. J. N. *et al.* 1998. Kirsten ras Mutations in Patients With Colorectal Cancer: the Multicenter "RASCAL" Study. *Journal of the National Cancer Institute* 90(9), pp. 675-684.

Andreyev, H. J. N. *et al.* 2001. Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study. *Br J Cancer* 85(5), pp. 692-696.

Arai, K. *et al.* 1997. Cloning of a human homolog of the yeast OGG1 gene that is involved in the repair of oxidative DNA damage. *Oncogene* 14(23), pp. 2857-2861.

Arcila, M. *et al.* 2011. Detection of KRAS and BRAF Mutations in Colorectal Carcinoma: Roles for High-Sensitivity Locked Nucleic Acid-PCR Sequencing and Broad-Spectrum Mass Spectrometry Genotyping. *The Journal of molecular diagnostics : JMD* 13(1), pp. 64-73.

Arnold, D. *et al.* 2008. Cetuximab in combination with weekly 5-fluorouracil/folinic acid and oxaliplatin (FUFOX) in untreated patients with advanced colorectal cancer: a phase Ib/II study of the AIO GI Group. *Annals of Oncology* 19(8), pp. 1442-1449.

Arnold, D. and Seufferlein, T. 2010. Targeted treatments in colorectal cancer: state of the art and future perspectives. *Gut* 59(6), pp. 838-858.

Arnould, S. *et al.* 2003. Cellular determinants of oxaliplatin sensitivity in colon cancer cell lines. *European journal of cancer (Oxford, England : 1990)* 39(1), pp. 112-119.

Asagoshi, K. *et al.* 2010. FEN1 Functions in Long Patch Base Excision Repair Under Conditions of Oxidative Stress in Vertebrate Cells. *Mol Cancer Res*, 8(2), pp. 204-215.

Aspinwall, R. *et al.* 1997. Cloning and characterization of a functional human homolog of *Escherichia coli* endonuclease III. *Proc Natl Acad Sci USA*, 94(1), pp. 109-114.

Augenlicht, L. H. *et al.* 1997. Low-Level c-*myc* Amplification in Human Colonic Carcinoma Cell Lines and Tumors: A Frequent, p53-independent Mutation Associated with Improved Outcome in a Randomized Multi-institutional Trial. *Cancer Research* 57, pp. 1769-1775.

Azzato, E. M. *et al.* 2010. A genome-wide association study of prognosis in breast cancer. *Cancer Epidemiol Biomarkers Prev*, 19(4), pp. 1140-1143.

Azzopardi, D. *et al.* 2008. Multiple Rare Nonsynonymous Variants in the Adenomatous Polyposis Coli Gene Predispose to Colorectal Adenomas. *Cancer Research* 68(2), pp. 358-363.

Bajetta, E. *et al.* 2004. Randomized multicenter Phase II trial of two different schedules of irinotecan combined with capecitabine as first-line treatment in metastatic colorectal carcinoma. *Cancer* 100(2), pp. 279-287.

Bambury, R. and McCaffrey, J. 2009. Trichomegaly of the Eyelashes After Colorectal Cancer Treatment With the Epidermal Growth Factor Receptor Inhibitor Cetuximab. *Clinical Colorectal Cancer* 8(4), pp. 235-235.

Bamford, S. *et al.* 2004. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 91(2), pp. 355-358.

Barault, L. *et al.* 2008. Mutations in the RAS-MAPK, PI(3)K (phosphatidylinositol-3-OH kinase) signaling network correlate with poor survival in a population-based series of colon cancers. *Int J Cancer* 122(10), pp. 2255 - 2259.

Bardi, G. *et al.* 1997. Cytogenetic comparisons of synchronous carcinomas and polyps in patients with colorectal cancer. *Br J Cancer* 76(6), pp. 765-769.

Bardou, V. *et al.* 2003. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *J Clin Oncol* 21, pp. 1973 - 1979.

Bardwell, P. D. *et al.* 2004. Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol*, 5(2), pp. 224-229.

Barrett, J. C. *et al.* 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2), pp. 263-265.

Baselga, J. and Rosen, N. 2008. Determinants of RASistance to anti-epidermal growth factor receptor agents. *J Clin Oncol* 26(10), pp. 1582 - 1584.

Batlle, E. *et al.* 2002. β-Catenin and TCF Mediate Cell Positioning in the Intestinal Epithelium by Controlling the Expression of EphB/EphrinB. *Cell* 111(2), pp. 251-263.

Baudis, M. and Cleary, M. 2001. Progenetix.net: an online repository for molecular cytogenetic aberration data. *Bioinformatics* 17(12), pp. 1228 - 1229.

Baylln, S. B. *et al.* 1997. Alterations in DNA Methylation: A Fundamental Aspect of Neoplasia. In: George, F.V.W. and George, K. eds. *Advances in Cancer Research*. Vol. Volume 72. Academic Press, pp. 141-182, 182a, 182b, 183-196.

Beckmann, J. S. and Weber, J. L. 1992. Survey of human and rat microsatellites. *Genomics* 12(4), pp. 627-631.

Becquemont, L. *et al.* 2011. Practical recommendations for pharmacogenomics-based prescription: 2010 ESF–UB Conference on Pharmacogenetics and Pharmacogenomics. *Pharmacogenomics* 12(1), pp. 113-124.

Bellacosa, A. *et al.* 1999. MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc Natl Acad Sci USA*, 96(7), pp. 3969-3974.

Bello-Fernandez, C. *et al.* 1993. The ornithine decarboxylase gene is a transcriptional target of c-Myc. *Proceedings of the National Academy of Sciences* 90(16), pp. 7804-7808.

Bellovin, D. I. *et al.* 2006. Reciprocal regulation of RhoA and RhoC characterizes the EMT and identifies RhoC as a prognostic marker of colon carcinoma. *Oncogene* 25(52), pp. 6959-6967.

Bensaad, K. *et al.* 2006. TIGAR, a p53-Inducible Regulator of Glycolysis and Apoptosis. *Cell*, 126(1), pp. 107-120.

Berardi, R. *et al.* 2010. Panitumumab: the evidence for its use in the treatment of metastatic colorectal cancer. *Core evidence* 5, pp. 61-76.

Berkel, H. J. *et al.* 2001. Expression of the Translation Initiation Factor eIF4E in the Polyp-Cancer Sequence in the Colon. *Cancer Ep Bio Prev* 10(6), pp. 663-666.

Berndt, S. I. *et al.* 2008. Pooled analysis of genetic variation at chromosome 8q24 and colorectal neoplasia risk. *Human Molecular Genetics* 17(17), pp. 2665-2672.

Bernoux, A. *et al.* 1998. Estrogen receptor negative and progesterone receptor positive primary breast cancer: pathological characteristics and clinical outcome. Institut Curie Breast Cancer Study Group. *Breast Cancer Res Treat* 49, pp. 219 - 225.

Béroud, C. and Soussi, T. 1996. APC Gene: Database of Germline and Somatic Mutations in Human Tumors and Cell Lines. *Nucleic Acids Research* 24(1), pp. 121-124.

Berry, D. *et al.* 2006. Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. *JAMA* 295, pp. 1658 - 1667.

Bindra, R. S. and Glazer, P. M. 2007. Co-repression of mismatch repair gene expression by hypoxia in cancer cells: Role of the Myc/Max network. *Cancer Lett*, 252(1), pp. 93-103.

Bjoras, M. *et al.* 1997. Opposite base-dependent reactions of a human base excision repair enzyme on DNA containing 7,8-dihydro-8-oxoguanine and abasic sites. *EMBO J* 16(20), pp. 6314-6322.

Black, P. C. *et al.* 2008. Sensitivity to epidermal growth factor receptor inhibitor requires E-cadherin expression in urothelial carcinoma cells. *Clin Cancer Res* 14(5), pp. 1478-1486.

Bland, J. M. and Altman, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1(8476), pp. 307-310.

Bleiberg, H. 1998. Oxaliplatin (L-OHP): a new reality in colorectal cancer. Nature Publishing Group.

Blons, H. *et al.* 1999. Frequent Allelic Loss at Chromosome 3p Distinct from Genetic Alterations of the 8-Oxoguanine DNA Glycosylase 1 Gene in Head and Neck Cancer. *Mol Carcinog*, 26(4), pp. 254-260.

Boland, C. R. and Goel, A. 2010. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138(6), pp. 2073-2087.e2073.

Boland, C. R. *et al.* 2009. Promoter Methylation in the Genesis of Gastrointestinal Cancer. *Yonsei Med J* 50(3), pp. 309-321.

Borner, M. M. *et al.* 2005. A randomized phase II trial of capecitabine and two different schedules of irinotecan in first-line treatment of metastatic colorectal cancer: efficacy, quality-of-life and toxicity. *Annals of Oncology* 16(2), pp. 282-288.

Borner, M. M. *et al.* 2006. The impact of cetuximab on the capecitabine plus oxaliplatin (XELOX) combination in firstline treatment of metastatic colorectal cancer: A randomized phase II trial of the Swiss Group for Clinical Cancer Research (SAKK). *J Clin Oncol. ASCO Annual Meeting Proceedings* 1(24), abstr 18S.

Bouali, S. *et al.* 2009. P53 and PTEN expression contribute to the inhibition of EGFR downstream signaling pathway by cetuximab. *Cancer Gene Ther* 16(6), pp. 498-507.

Braithwaite, E. K. *et al.* 2005. DNA polymerase lambda protects mouse fibroblasts against oxidative DNA damage and is recruited to sites of DNA damage/repair. *J Biol Chem*, 280(36), pp. 31641-31647.

Brata Das, B. *et al.* 2010. Role of tyrosyl-DNA phosphodiesterase (TDP1) in mitochondria. *Proc Natl Acad Sci USA*, 107(46), pp. 19790-19795.

Brink, M. *et al.* 2003. K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis* 24(4), pp. 703-710.

Broderick, P. *et al.* 2007. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet* 39(11), pp. 1315-1317.

Brugger, W. 2010. Successful treatment with the fully human antibody panitumumab after a severe infusion reaction with cetuximab. *Tumori* 96(3), pp. 473-477.

Bruns, C. J. *et al.* 2000. Epidermal Growth Factor Receptor Blockade with C225 Plus Gemcitabine Results in Regression of Human Pancreatic Carcinoma Growing Orthotopically in Nude Mice by Antiangiogenic Mechanisms. *Clinical Cancer Research* 6(5), pp. 1936-1948.

Bruns, C. J. *et al.* 2000. Blockade of the Epidermal Growth Factor Receptor Signaling by a Novel Tyrosine Kinase Inhibitor Leads to Apoptosis of Endothelial Cells and Therapy of Human Pancreatic Carcinoma. *Cancer Research* 60(11), pp. 2926-2935.

Burgess, A. W. *et al.* 2003. An Open-and-Shut Case? Recent Insights into the Activation of EGF/ErbB Receptors. *Molecular cell* 12(3), pp. 541-552.

Buyse, M. *et al.* 2010. Biomarkers and surrogate end points-the challenge of statistical validation. *Nat Rev Oncol* 7(6), pp. 309-317.

Cappelli, E. *et al.* 1997. Involvement of XRCC1 and DNA Ligase III Gene Products in DNA Base Excision Repair. *J Biol Chem*, 272(38), pp. 23970-23975.

Cardoso, J. *et al.* 2006. Chromosomal Instability in MYH- and APC-Mutant Adenomatous Polyps. *Cancer Research* 66(5), pp. 2514-2519.

Carpten, J. *et al.* 2007. A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* 448(7152), pp. 439 - 444.

Carvajal-Carmona, L. G. *et al.* 2011. The *GREM1, BMP4* and *BMP2* loci harbor multiple common susceptibility variants for colorectal cancer. *In submission.*

Carvajal-Carmona, L.G. *et al.* 2011. Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. *Human Mol Genet* 20(14), pp. 2879-2888.

Cassidy, J. *et al.* 2004. XELOX (Capecitabine Plus Oxaliplatin): Active First-Line Therapy for Patients With Metastatic Colorectal Cancer. *Journal of Clinical Oncology* 22(11), pp. 2084-2091.

Cassidy, J. *et al.* 2008. Efficacy of capecitabine vs. 5-FU in colorectal and gastric cancer: Meta-analysis of survival in 6 clinical trials. *Proc Am Soc Clin Oncol*, abstr 404.

Chan, A. T. and Giovannucci, E. L. 2010. Primary Prevention of Colorectal Cancer. *Gastroenterology* 138(6), pp. 2029-2043.e2010.

Cheadle, J. P. *et al.* 2002. Different Combinations of Biallelic APC Mutation Confer Different Growth Advantages in Colorectal Tumours. *Cancer Research* 62(2), pp. 363-366.

Cheadle, J. P. and Sampson, J. R. 2003. Exposing the MYtH about base excision repair and human inherited disease. *Human Molecular Genetics* 12(suppl 2), pp. R159-R165.

Cheadle, J. P. and Sampson, J. R. 2007. MUTYH-associated polyposis-from defect in base excision repair to clinical genetic testing. *DNA repair* 6(3), pp. 274-279.

Chen, R. *et al.* 2007. Identification of JWA as a novel functional gene responsive to environmental oxidative stress induced by benzo[a]pyrene and hydrogen peroxide. *Free radic Biol Med*, 42(11), pp. 1704-1714.

Chua, W. *et al.* 2009. Molecular markers of response and toxicity to FOLFOX chemotherapy in metastatic colorectal cancer. *Br J Cancer* 101(6), pp. 998-1004.

Chung, C. H. *et al.* 2008. Cetuximab-Induced Anaphylaxis and IgE Specific for Galactose-α-1,3-Galactose. *New England Journal of Medicine* 358(11), pp. 1109-1117.

Cianfrocca, M. and Goldstein, L. 2004. Prognostic and predictive factors in early-stage breast cancer. *Oncologist* 9, pp. 606 - 616.

Ciardiello, F. *et al.* 2000. Antiangiogenic and Antitumor Activity of Anti-Epidermal Growth Factor Receptor C225 Monoclonal Antibody in Combination with Vascular Endothelial Growth Factor Antisense Oligonucleotide in Human GEO Colon Cancer Cells. *Clinical Cancer Research* 6(9), pp. 3739-3747.

Cicek, M. S. *et al.* 2009. Functional and clinical significance of variants localized to 8q24 in colon cancer. *Cancer Epidemiol Biomarkers Prev*, 18(9), pp. 2492-2500.

Cleary, S. P. *et al.* 2009. Germline MutY Human Homologue Mutations and Colorectal Cancer: A Multisite Case-Control Study. *Gastroenterology*, 136(4), pp. 1251-1260.

Clevers, H. 2006. Wnt/β-Catenin Signaling in Development and Disease. *Cell* 127(3), pp. 469-480.

Collela, S. *et al.* 2003. Sensitive and quantitative universal Pyrosequencing methylation analysis. *Biotechniques* 35(1), pp. 146-150.

Collins, A. R. *et al.* 2004. Are we sure we know how to measure 8-oxo-7,8-dihydroguanine in DNA from human cells? *Archives of Biochemistry and Biophysics* 423(1), pp. 57-65.

Comella, P. *et al.* 2009. Role of oxaliplatin in the treatment of colorectal cancer. *Therapeutics and Clinical Risk Management* 5, pp. 229-238.

Conlin, A. *et al.* 2005. The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma. *Gut* 54(9), pp. 1283 - 1286.

Cooke, M. S. and Evans, M. D. 2007. 8-Oxo-deoxyguanosine: Reduce, reuse, recycle? *Proceedings of the National Academy of Sciences* 104(34), pp. 13535-13536.

Cortellino, S. *et al.* 2003. The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proc Natl Acad Sci USA*, 100(25), pp. 15071-15076.

Costa, S. *et al.* 2002. Factors influencing the prognostic role of oestrogen and progesterone receptor levels in breast cancer - results of the analysis of 670 patients with 11 years of follow-up. *Eur J Cancer* 38, pp. 1329 - 1334.

Cowper, D. *et al.* 2002. A primer and comparative review of major US mortality databases. *Ann Epidemiol* 12, pp. 462 - 468.

Cox, D. 1972. Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973-2001), National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program, Cancer Statistics Branch, released April based on the November 2003 submission

Regression models and life tables (with discussion). *J R Stat Soc Ser B* 34, pp. 187 -
220.

Crook, S. *et al.* 2009. Concomitant mutations and splice variants in KRAS and BRAF
demonstrate complex perturbation of the Ras/Raf signalling pathway in Colorectal
Cancer. *Gut.*

Crowe, J. *et al.* 1991. Estrogen receptor determination and long term survival of
patients with carcinoma of the breast. *Surg Gynecol Obstet* 173, pp. 273 - 278.

Culligan, K. M. *et al.* 2000. Evolutionary origin, diversification and specialization of
eukaryotic MutS homolog mismatch repair proteins. *Nucleic Acids Research* 28(2),
pp. 463-471.

Cunningham, D. *et al.* 2010. Colorectal cancer. *The Lancet* 375(9719), pp. 1030-
1047.

Cunningham, D. *et al.* 2004. Cetuximab Monotherapy and Cetuximab plus Irinotecan
in Irinotecan-Refractory Metastatic Colorectal Cancer. *New England Journal of
Medicine* 351(4), pp. 337-345.

Curtin, K. *et al.* 2009. Meta Association of Colorectal Cancer Confirms Risk Alleles at
8q24 and 18q21. *Cancer Epidemiology Biomarkers & Prevention* 18(2), pp. 616-621.

Dallosso, A. R. *et al.* 2008. Inherited predisposition to colorectal adenomas caused
by multiple rare alleles of MUTYH but not OGG1, NUDT1, NTH1 or NEIL 1, 2 or 3.
*Gut* 57(9), pp. 1252-1255.

Daniels, D. L. and Weis, W. I. 2005. [beta]-catenin directly displaces Groucho/TLE
repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat Struct Mol Biol*
12(4), pp. 364-371.

Das, S. *et al.* 2007. Stimulation of NEIL2-mediated oxidized base excision repair via
YB-1 interaction during oxidative stress. *J Biol Chem,* 282(39), pp. 28474-28484.

Davies, H. *et al.* 2002. Mutations of the BRAF gene in human cancer. *Nature* 417(6892), pp. 949-954.

De Benedetti, A. *et al.* 1994. CHO cells transformed by the translation factor eIF4E display increased c-myc expression, but require overexpression of Max for tumorigenicity. *Mol and Cell Diff* 2(4), pp. 347-371.

De Roosk, E. *et al.* 2008. KRAS-wild type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann Oncol* 19(3), pp. 508-515.

De Roock, W. *et al.* 2010a. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *The Lancet Oncology* 11(8), pp. 753-762.

De Roock, W. *et al.* 2010b. Association of KRAS p.G13D Mutation With Outcome in Patients With Chemotherapy-Refractory Metastatic Colorectal Cancer Treated With Cetuximab. *JAMA: The Journal of the American Medical Association* 304(16), pp. 1812-1820.

Dherin, C. *et al.* 1999. Excision of oxidatively damaged DNA bases by the human alpha-hOgg1 protein and the polymorphic alpha-hOgg1 (Ser 326Cys) protein which is frequently found in human populations. *Nucl Acids Res* 27(20), pp. 4001-4007.

Di Nicolantonio, F. *et al.* 2008. Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J Clin Oncol* 26(35), pp. 5705 - 5712.

Dianov, G. *et al.* 1999. Repair of 8-oxoguanine in DNA is deficient in Cockayne syndrome group B cells. *Nucleic Acids Res*, 27(5), pp. 1365-1368.

Díaz-Rubio, E. *et al.* 2002. Capecitabine (Xeloda®) in combination with oxaliplatin: a phase I, dose-escalation study in patients with advanced or metastatic solid tumors. *Annals of Oncology* 13(4), pp. 558-565.

Dillon, D. *et al.* 2001. Rapid, Efficient Genotyping of Clinical Tumor Samples by Laser-Capture Microdissection/PCR/SSCP. *Experimental and Molecular Pathology* 70(3), pp. 195-200.

Doherty, K. M. *et al.* 2005. RECQ1 Helicase Interacts with Human Mismatch Repair Factors That Regulate Genetic Recombination. *J Biol Chem*, 280(30), pp. 28085-28094.

Donegan, W. 1992. Prognostic factors. Stage and receptor status in breast cancer. *Cancer* 70(6 Suppl), pp. 1755 - 1764.

Doroshow, J. *et al.* 1990. Prospective randomized comparison of fluorouracil versus fluorouracil and high-dose continuous infusion leucovorin calcium for the treatment of advanced measurable colorectal cancer in patients previously unexposed to chemotherapy. *Journal of Clinical Oncology* 8(3), pp. 491-501.

Doroshow, N. P. *et al.* 1992. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. Advanced Colorectal Cancer Meta-Analysis Project. *Journal of Clinical Oncology* 10(6), pp. 896-903.

Dotor, E. *et al.* 2006. Tumor Thymidylate Synthase 1494del6 Genotype As a Prognostic Factor in Colorectal Cancer Patients Receiving Fluorouracil-Based Adjuvant Treatment. *Journal of Clinical Oncology* 24(10), pp. 1603-1611.

Douillard, J. Y. *et al.* 2000. Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet* 355(9209), pp. 1041-1047.

Dunnwald, L. *et al.* 2007. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Research* 9(1), p. R6.

Easton, D. F. *et al.* 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447(7148), pp. 1087-1093.

Elledge, R. *et al.* 1994. Tumor biologic factors and breast cancer prognosis among white, Hispanic, and black women in the United States. *J Natl Cancer Inst* 86, pp. 705 - 712.

Engelward, B. P. *et al.* 1997. Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase. *Proc Natl Acad Sci USA*, 94(24), pp. 13087-13092.

Fakhrai-Rad, H. *et al.* 2002. Pyrosequencing™: An accurate detection platform for single nucleotide polymorphisms. *Human Mutation* 19(5), pp. 479-485.

Fakih, M. and Vincent, M. 2010. Adverse events associated with anti-EGFR therapies for the treatment of metastatic colorectal cancer. *Current oncology (Toronto, Ont.)* 17 Suppl 1, pp. S18-30.

Fariña-Sarasqueta, A. *et al.* 2010. The BRAF V600E mutation is an independent prognostic factor for survival in stage II and stage III colon cancer patients. *Annals of Oncology.*

Farrington, S. M. *et al.* 2005. Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet* 77(1), pp. 112-119.

Fearnhead, N. S. *et al.* 2001. The ABC of APC. *Human Molecular Genetics* 10(7), pp. 721-733.

Fearnhead, N. S. *et al.* 2004. Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proceedings of the*

*National Academy of Sciences of the United States of America* 101(45), pp. 15992-15997.

Fearnhead, N.S. *et al.* Rare variant hypothesis for multifactorial inheritance: susceptibility to colorectal adenomas as a model. *Cell Cycle* 4(4), pp. 521-525. Fearon, E. R. and Vogelstein, B. 1990. A genetic model for colorectal tumourigenesis. *Cell* 61(5), pp. 759-767.

Fink, D. *et al.* 1997. In Vitro and in Vivo Resistance to Cisplatin in Cells That Have Lost DNA Mismatch Repair. *Cancer Research* 57(10), pp. 1841-1845.

Fishel, R. *et al.* 1993. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* 75(5), pp. 1027-1038.

Fisher, B. *et al.* 2004. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *Lancet* 364, pp. 858 - 868.

Fisher, B. *et al.* 1988. Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. *J Clin Oncol* 6, pp. 1076 - 1087.

Fleischmann, C. *et al.* 2004. Comprehensive analysis of the contribution of germline MYH variation to early-onset colorectal cancer. *International Journal of Cancer* 109(4), pp. 554-558.

Folprecht, G. *et al.* 2010. Tumour response and secondary resectability of colorectal liver metastases following neoadjuvant chemotherapy with cetuximab: the CELIM randomised phase 2 trial. *Lancet Oncol* 11(1), p38-47.

Forbes, S. *et al.* 2008. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* Chapter 10(Unit 10), p. 11.

Forrester, K. *et al.* 1987. Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature* 327(6120), pp. 298-303.

Fortini, P. *et al.* 2003. The base excision repair: mechanisms and its relevance for cancer susceptibility. *Biochimie*, 84(11), pp. 1053-1071.
Fortini, P. and Dogliotti, E. 2007. Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair* 6(4), pp. 398-409.

Fox *et al.* 2006. *Pathology and management of dermatologic toxicities associated with anti-EGFR therapy.* Norwalk, CT, ETATS-UNIS: UBM Medica, p. 9.

Frattini, M. *et al.* 2007. PTEN loss of expression predicts cetuximab efficacy in metastatic colorectal cancer patients. *Br J Cancer* 97(8), pp. 1136-1145.

Frayling, I. M. *et al.* 1998. The APC variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history. *Proceedings of the National Academy of Sciences* 95(18), pp. 10722-10727.

Freeman, D. *et al.* 2008. Association of K-ras Mutational Status and Clinical Outcomes in Patients with Metastatic Colorectal Cancer Receiving Panitumumab Alone. *Clinical Colorectal Cancer* 7(3), pp. 184-190.

Fumagalli, D. *et al.* 2010. A rapid, sensitive, reproducible and cost-effective method for mutation profiling of colon cancer and metastatic lymph nodes. *BMC Cancer* 10(1), p. 101.

Funke, S. *et al.* 2010. Genetic polymorphisms in GST genes and survival of colorectal cancer patients treated with chemotherapy. *Pharmacogenomics* 11(1), pp. 33-41.

Galon, J. *et al.* 2006. Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science* 313(5795), pp. 1960-1964.

Gamblin, T. *et al.* 2006. Microdissection-based allelotyping: a novel technique to determine the temporal sequence and biological aggressiveness of colorectal cancer. *Am Surg* 72(5), pp. 445 - 453.

Gapstur, S. *et al.* 1996. Hormone receptor status of breast tumors in black, Hispanic, and non-Hispanic white women. An analysis of 13,239 cases. *Cancer* 77, pp. 1465 - 1471.

Garinis, G. A. *et al.* 2002. Hypermethylation-associated transcriptional silencing of E-cadherin in primary sporadic colorectal cancer. *J Pathol* 198(4), pp. 442-449.

Ge Guo. *et al.* 2004. Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature*, 429(6994), pp. 891-895.

Giantonio, B. J. *et al.* 2007. Bevacizumab in Combination With Oxaliplatin, Fluorouracil, and Leucovorin (FOLFOX4) for Previously Treated Metastatic Colorectal Cancer: Results From the Eastern Cooperative Oncology Group Study E3200. *Journal of Clinical Oncology* 25(12), pp. 1539-1544.

Gibbs, R. A. (The Internation HapMap Consortium). 2003. The International HapMap Project. *Nature* 426(6968), pp. 789-796.

Gismondi, V. *et al.* 2002. Prevalence of the E1317Q Variant of the APC Gene in Italian Patients with Colorectal Adenomas. *Genetic Testing* 6(4), pp. 313-317.

Goel, A. *et al.* 2003. Characterization of Sporadic Colon Cancer by Patterns of Genomic Instability. *Cancer Research* 63(7), pp. 1608-1614.

Goldhirsch, A. *et al.* 2003. Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J Clin Oncol* 21, pp. 3357 - 3365.

Grady, W. M. and Markowitz, S. D. 2002. Genetic and epigenetic alterations in colon cancer. *Annual Review of Genomics and Human Genetics* 3(1), pp. 101-128.

Graff, J. R. *et al.* 2007. Therapeutic suppression of translation initiation factor eIF4E expression reduces tumor growth without toxicity. *J Clin Invest* 117(9), pp. 2638-2648.

Grann, V. *et al.* 2005. Hormone receptor status and survival in a population-based cohort of patients with breast carcinoma. *Cancer* 103, pp. 2241 - 2251.

Griffiths, G. J. *et al.* 2004. Expression of kinase-defective mutants of c-Src in human metastatic colon cancer cells decreases Bcl-x(L) and increases oxaliplatin and Fas-induced apoptosis. *Journal of biological chemistry.* 279(44), pp. 46113-46121.

Groden, J. *et al.* 1991. Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* 66(3), pp. 589-600.

Grothey, A. 2006. Recognizing and managing toxicities of molecular targeted therapies for colorectal cancer. *Oncology (Willston Park)* 20(14 Suppl 10), pp. 21-28.

Group, I. B. C. S. 2002. *Endocrine responsiveness and tailoring adjuvant therapy for postmenopausal lymph node-negative breast cancer: a randomized trial.* Lyon, France: IARC Press, pp. 1054 - 1065.

Gruber, S. B. 2007. Genetic variation in 8q24 associated with risk of colorectal cancer. *Cancer Biology and Therapy* 6(7), pp. 1143-1147.

Gryfe, R. *et al.* 1999. Inherited colorectal polyposis and cancer risk of the APC I1307K polymorphism. *American journal of human genetics* 64(2), pp. 378-384.

Hah, S. S. *et al.* 2007. Measurement of 7,8-dihydro-8-oxo-2'-deoxyguanosine metabolism in MCF-7 cells at low concentrations using accelerator mass spectrometry. *Proceedings of the National Academy of Sciences* 104(27), pp. 11203-11208.

Hahnloser, D. *et al.* 2003. The APC E1317Q Variant in Adenomatous Polyps and Colorectal Cancers. *Cancer Epidemiology Biomarkers & Prevention* 12(10), pp. 1023-1028.

Haiman, C. A. *et al.* 2007. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39(5), pp. 638-644.

Half, E. *et al.* 2009. Familial adenomatous polyposis. *Orphanet Journal of Rare Diseases* 4(1), p. 22.

Halling, K. C. *et al.* 1999. Microsatellite Instability and 8p Allelic Imbalance in Stage B2 and C Colorectal Cancers. *Journal of the National Cancer Institute* 91(15), pp. 1295-1303.

Hampel, H. *et al.* 2008. Feasibility of Screening for Lynch Syndrome Among Patients With Colorectal Cancer. *Journal of Clinical Oncology* 26(35), pp. 5783-5788.

Harvey, J. *et al.* 1999. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* 17, pp. 1474 - 1481.

Hashigasako, A. *et al.* 2004. Bi-directional Regulation of Ser-985 Phosphorylation of c-Met via Protein Kinase C and Protein Phosphatase 2A Involves c-Met Activation and Cellular Responsiveness to Hepatocyte Growth Factor. *J Biol Chem* 279(25), pp. 26445 - 26452.

Haushalter, K. A. *et al.* 1999. Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr Biol*, 9(4), pp. 174-185.

Haydon, A. M. M. *et al.* 2006. Effect of physical activity and body size on survival after diagnosis with colorectal cancer. *Gut* 55(1), pp. 62-67.

Hayes, V. M. *et al.* 2000. New comprehensive denaturing-gradient-gel-electrophoresis assay for KRAS mutation detection applied to paraffin-embedded tumours. *Genes, Chromosomes and Cancer* 29(4), pp. 309-314.

Hazra, T. K. *et al.* 2007. Oxidative DNA damage repair in mammalian cells: A new perspective. *DNA Repair* 6(4), pp. 470-480.

Hecht, J. R. *et al.* 2009. A Randomized Phase IIIB Trial of Chemotherapy, Bevacizumab, and Panitumumab Compared With Chemotherapy and Bevacizumab Alone for Metastatic Colorectal Cancer. *Journal of Clinical Oncology* 27(5), pp. 672-680.

Heidelberger, C. *et al.* 1957. Fluorinated Pyrimidines, A New Class of Tumour-Inhibitory Compounds. *Nature* 179(4561), pp. 663-666.

Hirano, T. 2008. Repair System of 7, 8-Dihydro-8-Oxoguanine as a Defense Line against Carcinogenesis. *Journal of Radiation Research* 49(4), pp. 329-340.

Hodgson, S.V and Maher, E.R. 1999. A practical guide to human cancer genetics. 1st ed. Cambridge: Cambridge University Press.

Hoeijmakers, J. H. J. 2001. Genome maintenance mechanisms for preventing cancer. *Nature* 411(6835), pp. 366-374.

Hollingsworth, N. M. *et al.* 1995. MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair. *Genes Dev*, 9(14), pp. 1728-1739.

Hollstein, M. *et al.* 1991. p53 mutations in human cancers. *Science* 253(5015), pp. 49-53.

Hou, Y. *et al.* 2007. Bcl2 impedes DNA mismatch repair by directly regulating the hMSH2-hMSH6 heterodimeric complex. *J Biol Chem*, 282(12), pp. 9279-9287.

Houlston, R. S. (COGENT study) *et al.* 2008. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 40(12), pp. 1426-1435.

Houlston, R. S. *et al.* 2010. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet* 42(11), pp. 973-977.

Hsieh, A. C. and Ruggero, D. 2010. Targeting eukaryotic translation initiation factor 4E (eIF4E) in Cancer. *Clin Cancer Res* 16(20), pp. 4914-4920.

Huang, S.-M. *et al.* 2002. Molecular Inhibition of Angiogenesis and Metastatic Potential in Human Squamous Cell Carcinomas after Epidermal Growth Factor Receptor Blockade 1 Supported in part by a research award from ImClone Systems Inc.1. *Molecular Cancer Therapeutics* 1(7), pp. 507-514.

Hubbard, S. R. 2005. EGF receptor inhibition: Attacks on multiple fronts. *Cancer cell* 7(4), pp. 287-288.

Hurwitz, H. *et al.* 2004. Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for Metastatic Colorectal Cancer. *New England Journal of Medicine* 350(23), pp. 2335-2342.

Hutter, C. *et al.* 2010. Characterization of the association between 8q24 and colon cancer: gene-environment exploration and meta-analysis. *BMC Cancer* 10(1), p. 670.

Hynes, N. E. and Lane, H. A. 2005. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 5(5), pp. 341-354.

Ide, Y. *et al.* 2003. Characterization of the genomic structure and expression of the mouse Apex2 gene. *Genomics* 81(1), pp. 47-57.

Ihalainen, J. *et al.* 1994. Diagnosis of pancreatic lesions using fine needle aspiration cytology: detection of K-ras point mutations using solid phase minisequencing. *Journal of Clinical Pathology* 47(12), pp. 1082-1084.

Ince, W. L. *et al.* 2005. Association of k-ras, b-raf, and p53 Status With the Treatment Effect of Bevacizumab. *Journal of the National Cancer Institute* 97(13), pp. 981-989.

Interthal, H. *et al.* 2005. Human Tdp1 Cleaves a Broad Spectrum of Substrates Including Phosphoamide Linkages. *J Biol Chem*, 280(43), pp. 36518-36528.

Ionov, Y. *et al.* 1993. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 363(6429), pp. 558-561.

Issa, J.-P. 2004. CpG island methylator phenotype in cancer. *Nat Rev Cancer* 4(12), pp. 988-993.

Izzi, B. *et al.* 2010. A new approach to imprinting mutation detection in GNAS by Sequenom EpiTYPER system. *Clinica Chimica Acta* 411(23-24), pp. 2033-2039.

Jaeger, E. *et al.* 2008. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet* 40(1), pp. 26-28.

Jaruga, P. *et al.* 1994. Oxidative DNA base damage and antioxidant enzyme activities in human lung cancer. *FEBS Letters* 341(1), pp. 59-64.

Jascur, T. and Boland, C. R. 2006. Structure and function of the components of the human DNA mismatch repair system. *International Journal of Cancer* 119(9), pp. 2030-2035.

Jeyaprakash, A. *et al.* 1994. Mutagenesis of yeast MW104-1B strain has identified the uncharacterized PMS6 DNA mismatch repair gene locus and additional alleles of existing PMS1, PMS2 and MSH2 genes. *Mutat Res*, 325(1), pp. 21-29.

Jhawer, M. *et al.* 2008. PIK3CA Mutation/PTEN Expression Status Predicts Response of Colon Cancer Cells to the Epidermal Growth Factor Receptor Inhibitor Cetuximab. *Cancer Res* 68(6), pp. 1953-1961.

Jiang, M.-R. *et al.* 2003. c-Myc degradation induced by DNA damage results in apoptosis of CHO cells. *Oncogene* 22(21), pp. 3252-3259.

Jiricny, J. 2006. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol* 7(5), pp. 335-346.

Johnson, L. *et al.* 1997. K-ras is an essential gene in the mouse with partial functional overlap with N-ras. *Genes Dev* 11(19), pp. 2468-2481.

Jones, S. *et al.* 2002. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C→T:A mutations. *Human Molecular Genetics* 11(23), pp. 2961-2967.

Jonker, D. J. *et al.* 2007. Cetuximab for the Treatment of Colorectal Cancer. *N Engl J Med* 357(20), pp. 2040-2048

Jung, Y. D. *et al.* 2002. Effects of combination anti-vascular endothelial growth factor receptor and anti-epidermal growth factor receptor therapies on the growth of gastric cancer in a nude mouse model. *European journal of cancer (Oxford, England : 1990)* 38(8), pp. 1133-1140.

Kabbinavar, F. *et al.* 2003. Phase II, Randomized Trial Comparing Bevacizumab Plus Fluorouracil (FU)/Leucovorin (LV) With FU/LV Alone in Patients With Metastatic Colorectal Cancer. *Journal of Clinical Oncology* 21(1), pp. 60-65.

Kabbinavar, F. F. *et al.* 2005. Addition of Bevacizumab to Bolus Fluorouracil and Leucovorin in First-Line Metastatic Colorectal Cancer: Results of a Randomized Phase II Trial. *Journal of Clinical Oncology* 23(16), pp. 3697-3705.

Kalbfleisch, J. and Prentice, R. 2002. The statistical analysis of failure time data. 2nd ed. New York: John Wiley & Sons.

Karapetis, C. S. *et al.* 2008. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 359(17), pp. 1757-1765.

Kato, S. *et al.* 2007. PIK3CA mutation is predictive of poor survival in patients with colorectal cancer. *International Journal of Cancer* 121(8), pp. 1771-1778.

Kavli, B. *et al.* 2002. hUNG2 Is the Major Repair Enzyme for Removal of Uracil from U:A Matches, U:G Mismatches, and U in Single-stranded DNA, with hSMUG1 as a Broad Specificity Backup. *J Biol Chem*, 277(42), pp. 39926-39936.

Kawaguchi, Y. *et al.* 2007. Cetuximab induce antibody-dependent cellular cytotoxicity against EGFR-expressing esophageal squamous cell carcinoma. *International Journal of Cancer* 120(4), pp. 781-787.

Kentis, A. *et al.* 2004. Ribavirin suppresses eIF4E mediated oncogenic transformation. *Proc Natl Acad Sci USA* 101(52), pp. 18105-18110.

Khambata-Ford, S. *et al.* 2007. Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J Clin Oncol* 25(22), pp. 3230 - 3237.

Kheirelseid, E. A. H. *et al.* 2010. Identification of endogenous control genes for normalisation of real-time quantitative PCR data in colorectal cancer. *BMC Mol Biol* 11(12).

Kiemeney, L. A. *et al.* 2008. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet* 40(11), pp. 1307-1312.

Kim, B. H. *et al.* 2007. On the functions of the h subunit of eukaryotic initiation factor 3 in late stages of translation initiation. *Genome Biol* 8(4), pp. R60.

Kim, J. G. *et al.* 2008. Vascular Endothelial Growth Factor Gene Polymorphisms Associated with Prognosis for Patients with Colorectal Cancer. *Clinical Cancer Research* 14(1), pp. 62-66.

Kim, R. 2009. Cetuximab and panitumumab: are they interchangeable? *The Lancet Oncology* 10(12), pp. 1140-1141.

Kim, S. M. *et al.* 2010. Acquired resistance to cetuximab is mediated by increased PTEN instability and leads cross-resistance to gefitinib in HCC827 NSCLC cells. *Cancer Lett* 296(2), pp. 150-159.

Kimura, Y. *et al.* 2004. A variant form of *hMTH1*, a human homologue of the E coli mutT gene, correlates with somatic mutation in the p53 tumour suppressor gene in gastric cancer patients. *J Med Genet*, 41(5), e57.

Kinzler, K. W. and Vogelstein, B. 1996. Lessons from Hereditary Colorectal Cancer. *Cell* 87(2), pp. 159-170.

Kiyota, A. *et al.* 2002. Anti-Epidermal Growth Factor Receptor Monoclonal Antibody 225 Upregulates p27KIP1 and p15INK4B and Induces G1 Arrest in Oral Squamous Carcinoma Cell Lines. *Oncology* 63(1), pp. 92-98.

Klaunig, J. E. and Kamendulis, L. M. 2004. The role of oxidative stress in carcinogenesis. *Annual Review of Pharmacology and Toxicology* 44(1), pp. 239-267.

Klymkowsky, M. W. and Savagner, P. 2009. Epithelial-mesenchymal transition: a cancer researcher's conceptual friend and foe. *Am J Pathol* 174(5), pp. 1588-1593.

Knudson, A. G. 1996. Hereditary cancer: Two hits revisited. *Journal of Cancer Research and Clinical Oncology* 122(3), pp. 135-140.

Koessler, T. *et al.* 2009. Common germline variation in mismatch repair genes and survival after a diagnosis of colorectal cancer. *International Journal of Cancer* 124(8), pp. 1887-1891.

Kohno, T. *et al.* 1998. Genetic polymorphisms and alternative splicing of the hOGG1 gene, that is involved in the repair of 8-hydroxyguanine in damaged DNA. *Oncogene* 16(25), pp. 3219-3225.

Kohno, T. *et al.* 2006. Association of polymorphisms in the *MTH1* gene with small cell lung carcinoma risk. *Carcinogenesis*, 27(12), pp. 2448-2454.

Kolonel, L. N. *et al.* 2004. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nat Rev Cancer* 4(7), pp. 519-527.

Koo, J. S. *et al.* 2009. The predictive role of E-cadherin and androgen receptor on *in vitro* chemosensitivity in triple-negative breast cancer. *Jpn J Clin Oncol*, 39(9), pp. 560-568.

Kozak, M. 1991. An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol* 115(4), p. 887-903.

Kuebler, J. *et al.* 2007. Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: results from NSABP C-07. *J Clin Oncol* 25(16), pp. 2198 - 2204.

Kuismanen, S. A. *et al.* 2000. Genetic and Epigenetic Modification of MLH1 Accounts for a Major Share of Microsatellite-Unstable Colorectal Cancers. *The American Journal of Pathology* 156(5), pp. 1773-1779.

Kurzawski, G. *et al.* 2006. Germline MSH2 and MLH1 mutational spectrum including large rearrangements in HNPCC families from Poland (update study). *Clinical Genetics* 69(1), pp. 40-47.

Laken, S. J. *et al.* 1999. Analysis of masked mutations in familial adenomatous polyposis. *Proceedings of the National Academy of Sciences* 96(5), pp. 2322-2326.

Lambrechts, D. *et al.* 2009. The role of KRAS, BRAF, NRAS, and PIK3CA mutations as markers of resistance to cetuximab in chemorefractory metastatic colorectal cancer. *J Clin Oncol (Meeting Abstracts)* 27(15S), p. 4020.

Lamlum, H. *et al.* 2000. Germline APC variants in patients with multiple colorectal adenomas, with evidence for the particular importance of E1317Q. *Human Molecular Genetics* 9(15), pp. 2215-2221.

Laurent-Puig, P. *et al.* 2009. Analysis of PTEN, BRAF, and EGFR Status in Determining Benefit From Cetuximab Therapy in Wild-Type KRAS Metastatic Colon Cancer. *J Clin Oncol* 27(35), pp. 5924-5930.

Lazaris-Karatzas, A. *et al.* 1990. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature* 345(6275), pp. 544-547.

Leichmann, L. P. 2006. Fluoropyrimidine therapy: a new chapter in an old story. *Community Translation* 3(3), pp. 127-131.

Leitch, E. F. *et al.* 2007. Comparison of the prognostic value of selected markers of the systemic inflammatory response in patients with colorectal cancer. *Br J Cancer* 97(9), pp. 1266-1270.

Lethaby, A. *et al.* 1996. Survival of women with node negative breast cancer in the Auckland region. *N Z Med J* 109, pp. 330 - 333.

Levin, D. S. *et al.* 2000. Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Curr Biol*, 10(15), pp. 919-922.

Li, B. D. *et al.* 1997. Overexpression of eukaryotic initiation factor 4E (eIF4E) in breast carcinoma. *Cancer* 79(12), pp. 2385-2390.

Li, B. D. *et al.* 1998. Clinical outcome in stage I to III breast carcinoma and eIF4E overexpression. *Ann Surg* 227(5), pp. 756-761.

Li, L. C. *et al.* 2000. A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities. *Cancer Res*, 60(4), pp. 873-876.

Li, C. *et al.* 2003. Incidence of invasive breast cancer by hormone receptor status from 1992 to 1998. *J Clin Oncol* 21, pp. 28 - 34.

Li, C. *et al.* 2002. Differences in breast cancer hormone receptor status and histology by race and ethnicity among women 50 years of age and older. *Cancer Epidemiol Biomarkers Prev* 11, pp. 601 - 607.

Li, C. *et al.* 2005. National Cancer Comprehensive Network (NCCN) Clinical Practice Guidelines in Oncology, Breast Cancer, V.I.2007
Clinical characteristics of different histologic types of breast cancer. *Br J Cancer* 93, pp. 1046 - 1052.

Li, L. *et al.* 2008. A Common 8q24 Variant and the Risk of Colon Cancer: A Population-Based Case-Control Study. *Cancer Epidemiology Biomarkers & Prevention* 17(2), pp. 339-342.

Li, W. *et al.* 2006. BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer independently of microsatellite instability status. *Mol Cancer* 5, p. 2.

Lichtenstein, P. *et al.* 2000. Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *New England Journal of Medicine* 343(2), pp. 78-85.

Lievre, A. *et al.* 2006. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res* 66(8), pp. 3992-3995.

Lievre, A. *et al.* 2008. KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J Clin Oncol* 26(3), pp. 374 - 379.

Lin, Y. L. *et al.* 1998. The Evolutionarily Conserved Zinc Finger Motif in the Largest Subunit of Human Replication Protein A Is Required for DNA Replication and Mismatch Repair but Not for Nucleotide Excision Repair. *J Biol Chem*, 273(2), pp. 1453-1461.

Linardou, H. *et al.* 2008. Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. *The Lancet Oncology* 9(10), pp. 962 - 972.

Lindblom, A. *et al.* 1993. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet* 5(3), pp. 279-282.

Liu, B. *et al.* 2000. Induction of apoptosis and activation of the caspase cascade by anti-EGF receptor monoclonal antibodies in DiFi human colon cancer cells do not involve the c- jun N-terminal kinase activity. *Br J Cancer* 82(12), pp. 1991-1999.

Liu, G. *et al.* 1997. Patient preferences for oral versus intravenous palliative chemotherapy. *Journal of Clinical Oncology* 15(1), pp. 110-115.

Liu, L. F. *et al.* 2000. Mechanism of Action of Camptothecin. *Annals of the New York Academy of Sciences* 922(1), pp. 1-10.

Liu, M. *et al.* 2010. The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. *Proc Natl Acad Sci USA*, 107(11), pp. 4925-4930.

Longley, M. J. *et al.* 1997. DNA Polymerase δ Is Required for Human Mismatch Repair *in Vitro*. *J Biol Chem*, 272(16), pp. 10917-10921.

Longley, M. J. *et al.* 1998. Identification of 5-deoxyribose phosphate lyase activity in human DNA polymerase γ and its role in mitochondrial base excision repair in vitro. *Proc Natl Acad Sci USA*, 95(21), pp. 12244-12248.

Lorenzato, A. *et al.* 2002. Novel Somatic Mutations of the MET Oncogene in Human Carcinoma Metastases Activating Cell Motility and Invasion. *Cancer Res* 62(23), pp. 7025 - 7030.

Loriaux, M. *et al.* 2008. High-throughput sequence analysis of the tyrosine kinome in acute myeloid leukemia. *Blood* 111(9), pp. 4788 - 4796.

Loupakis, F. *et al.* 2009. PTEN Expression and KRAS Mutations on Primary Tumors and Metastases in the Prediction of Benefit From Cetuximab Plus Irinotecan for Patients With Metastatic Colorectal Cancer. *J Clin Oncol* 27(16), pp. 2622-2629.

Løvig, T. *et al.* 2002. APC and CTNNB1 Mutations in a Large Series of Sporadic Colorectal Carcinomas Stratified by the Microsatellite Instability Status. *Scand J Gastroenterol* 37(10), pp. 1184-1193.

Lu, A. L. *et al.* 2001. Repair of oxidative DNA damage. *Cell Biochemistry and Biophysics* 35(2), pp. 141-170.

Lubbe, S. J. *et al.* 2009. Clinical Implications of the Colorectal Cancer Risk Associated With MUTYH Mutation. *J Clin Oncol*, 27(24), pp. 3975-3980.

Lynch, H. T. *et al.* 2006. Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. *Eur J Hum Genet* 14(4), pp. 390-402.

Lynch, H. T. *et al.* 2007. Who Should Be Sent for Genetic Testing in Hereditary Colorectal Cancer Syndromes? *Journal of Clinical Oncology* 25(23), pp. 3534-3542.

Lynch, H. T. *et al.* 2009. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clinical Genetics* 76(1), pp. 1-18.

Ma, P. *et al.* 2003. c-MET Mutational Analysis in Small Cell Lung Cancer: Novel Juxtamembrane Domain Mutations Regulating Cytoskeletal Functions. *Cancer Res* 63(19), pp. 6272 - 6281.

Maher, E. R. *et al.* 1992. Phenotypic variation in hereditary nonpolyposis colon cancer syndrome association with infiltrative fibromatosis (desmoid tumor). *Cancer* 69(8), pp. 2049-2051.

Malins, D. C. and Haimanot, R. 1991. Major Alterations in the Nucleotide Structure of DNA in Cancer of the Female Breast. *Cancer Research* 51(19), pp. 5430-5432.

Mandal, M. *et al.* 1998. Nuclear targeting of Bax during apoptosis in human colorectal cancer cells. *Oncogene* 17(8), pp. 999-1007.

Mao, G. *et al.* 2007. Identification and characterization of *OGG1* mutations in patients with Alzheimer's disease. *Nucleic Acids Res*, 35(8), pp. 2759-2766.

Marcuello, E. *et al.* 2004. Single nucleotide polymorphism in the 5' tandem repeat sequences of thymidylate synthase gene predicts for response to fluorouracil-based chemotherapy in advanced colorectal cancer patients. *International Journal of Cancer* 112(5), pp. 733-737.

Markowitz, S. *et al.* 1995. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* 268(5215), pp. 1336-1338.

Marsin, S. *et al.* 2003. Role of XRCC1 in the coordination and stimulation of oxidative DNA damage repair initiated by the DNA glycosylase hOGG1. *J Biol Chem*, 278(45), pp. 44068-44074.

Martin, S. A. *et al.* 2010. DNA polymerases as potential therapeutic targets for cancers deficient in the DNA mismatch repair proteins MSH2 or MLH1. *Cancer Cell*, 17(3), pp. 235-248.

Martinelli, E. *et al.* 2010. Synergistic Antitumor Activity of Sorafenib in Combination with Epidermal Growth Factor Receptor Inhibitors in Colorectal and Lung Cancer Cells. *Clinical Cancer Research*.

Martin-Martorell, P. *et al.* 2008. Biweekly cetuximab and irinotecan in advanced colorectal cancer patients progressing after at least one previous line of chemotherapy: results of a phase II single institution trial. *Br J Cancer* 99(3), pp. 455-458.

Mascarenhas *et al.* 2009. New mutations detected by denaturing high performance liquid chromatography during screening of exon 6 bcr-abl mutations in patients with chronic myeloid leukemia treated with tyrosine kinase inhibitors. *Leuk* Lymphoma 50(7), pp. 1148-1154.

Masih, P. J. *et al.* 2008. Mismatch Repair proteins are recruited to replicating DNA through interaction with Proliferating Cell Nuclear Antigen (PCNA). *Nucleic Acids Res*, 36(1), pp. 67-75.

Matsubara, M. *et al.* 2004. Mutational analysis of the damage-recognition and catalytic mechanism of human SMUG1 DNA glycosylase. *Nucleis Acids Res*, 32(17), pp. 5291-5302.

Maughan, T. S. *et al.* 2002. British MRC Colorectal Cancer Working Party. Comparison of survival, palliation, and quality of life with three chemotherapy regimens in metastatic colorectal cancer: a multicentre randomised trial. *Lancet* 359(9317), pp. 1555-1563.

Maughan, T. S. *et al.* 2011. The addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for advanced colorectal cancer: results of the MRC COIN trial. *Lancet* in press.

Mayer, A. *et al.* 1993. The prognostic significance of proliferating cell nuclear antigen, epidermal growth factor receptor, and mdr gene expression in colorectal cancer. *Cancer* 71(8), pp. 2454-2460.

McLeod, H. L. *et al.* 1998. Characterization of dihydropyrimidine dehydrogenase in human colorectal tumours. *British journal of cancer* 77(3), pp. 461-465.

Ménessier-de Murcia, J. *et al.* 1997. Requirement of poly(ADP-ribose) polymerase in recovery from DNA damage in mice and in cells. *Proc Natl Acad Sci USA*, 94(14), pp. 7303-7307.

Menko, F. H. 1993. General aspects of colorectal cancer. In: Genetics of colorectal cancer for clinical practice. Kluwer academic publishers

Metges, J. *et al.* 2010. PANERB study: Panitumumab after cetuximab-based regimen failure. *Proc Am Soc Clin Oncol* 28, abstr e14000.

Metzger, R. *et al.* 1998. ERCC1 mRNA levels complement thymidylate synthase mRNA levels in predicting response and survival for gastric cancer patients receiving combination cisplatin and fluorouracil chemotherapy. *Journal of Clinical Oncology* 16(1), pp. 309-316.

Michaels, M. L. and Miller, J. H. 1992. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.* 174(20), pp. 6321-6325.

Michils, G. *et al.* 2005. Large deletions of the APC gene in 15% of mutation-negative patients with classical polyposis (FAP): A Belgian study. *Human Mutation* 25(2), pp. 125-134.

Miwa, M. *et al.* 1998. Design of a novel oral fluoropyrimidine carbamate, capecitabine, which generates 5-fluorouracil selectively in tumours by enzymes concentrated in human liver and cancer tissue. *European journal of cancer (Oxford, England : 1990)* 34(8), pp. 1274-1281.

Miyaki, M. *et al.* 1997. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet* 17(3), pp. 271-272.

Moerke, N. J. *et al.* 2007. Small-molecule inhibition of the interaction between the translational. *Cell* 128(2), pp. 257-267.

Moll, U. M. *et al.* 1992. Two distinct mechanisms alter p53 in breast cancer: mutation and nuclear exclusion. *Proceedings of the National Academy of Sciences* 89(15), pp. 7262-7266.

Mollevi, D. G. *et al.* 2008. PRL-3 is essentially overexpressed in primary colorectal tumours and associates with tumour aggressiveness. *Br J Cancer* 99(10), pp. 1718-1725.

Moosmann, N. *et al.* 2011. Cetuximab plus capecitabine and irinotecan compared with cetuximab plus capecitabine and oxaliplatin as first-line treatment for patients with metastatic colorectal cancer: AIO KRK-0104 – A randomized trial of the German AIO CRC study group. *J Clin Oncol*, 29(8), pp. 1050-1058.

Moreno, V. *et al.* 2006. Polymorphisms in Genes of Nucleotide and Base Excision Repair: Risk and Prognosis of Colorectal Cancer. *Clinical Cancer Research* 12(7), pp. 2101-2108.

Morin, P. J. *et al.* 1997. Activation of β-Catenin-Tcf Signaling in Colon Cancer by Mutations in β-Catenin or APC. *Science* 275(5307), pp. 1787-1790.

Morland, I. *et al.* 2002. Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res*, 30(22), pp. 4926-4936.

Moroni, M. *et al.* 2005. Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: a cohort study. *The Lancet Oncology* 6(5), pp. 279-286.

Morris, E. J. and Geller, H. M. 1996. Induction of neuronal apoptosis by camptothecin, an inhibitor of DNA topoisomerase-I: evidence for cell cycle-independent toxicity. *The Journal of Cell Biology* 134(3), pp. 757-770.

Mullis, K. *et al.* 1986. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 51, pp. 263-273.

Munro, A. J. *et al.* 2005. P53 abnormalities and outcomes in colorectal cancer: a systematic review. *Br J Cancer* 92(3), pp. 434-444.

Nagasaka, T. *et al.* 2008. Mutations in Both KRAS and BRAF May Contribute to the Methylator Phenotype in Colon Cancer. *Gastroenterology* 134(7), pp. 1950-1960.e1951.

Nakabeppu, Y. 2001. Molecular genetics and structural biology of human MutT homolog, MTH1. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 477(1-2), pp. 59-70.

Nathan, C. O. *et al.* 1997. Detection of the proto-oncogene eIF4E in surgical margins may predict recurrence in head and neck cancer. *Oncogene* 15(5), pp. 579-584.

Neddermann, P. and Jiricny, J. 1993. The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J Biol Chem*, 268(28), pp. 21218-21224.

Negri, F.V. *et al.* 2010. PTEN status in advanced colorectal cancer treated with cetuximab. *Br J Cancer* 102(1), pp. 162-164.

Negrini, S. *et al.* 2010. Genomic instability [mdash] an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 11(3), pp. 220-228.

Nelson, M. R. *et al.* 2004. Large-Scale Validation of Single Nucleotide Polymorphisms in Gene Regions. *Genome Research* 14(8), pp. 1664-1668.

Nicolaides, N. C. *et al.* 1994. Mutations of two P/WS homologues in hereditary nonpolyposis colon cancer. *Nature* 371(6492), pp. 75-80.

Nielsen, M. *et al.* 2010. Survival of *MUTYH*-Associated Polyposis Patients With Colorectal Cancer and Matched Control Colorectal Cancer Patients. *J Natl Cancer Inst*, 102(22), pp. 1724-1730.

Nikolova, D. A. *et al.* 2009. Cetuximab attenuates metastasis and u-PAR expression in non-small cell lung cancer: u-PAR and E-cadherin are novel biomarkers of cetuximab sensitivity. *Cancer Res* 69(6), pp. 2461-2470.

O'Neil, B. H. and Goldberg, R. M. 2005. Chemotherapy for advanced colorectal cancer: Let's not forget how we got here (until we really can). *Seminars in oncology* 32(1), pp. 35-42.

O-charoenrat, P. *et al.* 1999. Differential modulation of proliferation, matrix metalloproteinase expression and invasion of human head and neck squamous carcinoma cells by c-erbB ligands. *Clinical and Experimental Metastasis* 17(7), pp. 631-639.

O-Charoenrat, P. *et al.* 2000. Overexpression of epidermal growth factor receptor in human head and neck squamous carcinoma cell lines correlates with matrix metalloproteinase-9 expression and in vitro invasion. *International Journal of Cancer* 86(3), pp. 307-317.

Ogino, S. *et al.* 2005. Sensitive Sequencing Method for KRAS Mutation Detection by Pyrosequencing. *J Mol Diagn* 7(3), pp. 413-421.

Ogino, S. *et al.* 2009. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* 58(1), pp. 90 - 96.

Ogino, S. *et al.* 2009. PIK3CA Mutation Is Associated With Poor Prognosis Among Patients With Curatively Resected Colon Cancer. *J Clin Oncol.*

Ohtsubo, T. *et al.* 2000. Identification of human MutY homolog (hMYH) as a repair enzyme for 2-hydroxyadenine in DNA and detection of multiple forms of hMYH located in nuclei and mitochondria. *Nucleic Acids Research* 28(6), pp. 1355-1364.

Okamoto, K. *et al.* 1994. Formation of 8-hydroxy-2'-deoxyguanosine and 4-hydroxy-2-nonenal-modified proteins in human renal-cell carcinoma. *International Journal of Cancer* 58(6), pp. 825-829.

Okayama, N. *et al.* 2011. The Importance of Evaluation of DNA Amplificability in KRAS Mutation Testing with Dideoxy Sequencing using Formalin-fixed and Paraffin-

embedded Colorectal Cancer Tissues. *Japanese Journal of Clinical Oncology* 41(2), pp. 165-171.

Oldenhuis, C. N. A. M. *et al.* 2008. Prognostic versus predictive value of biomarkers in oncology. *European journal of cancer (Oxford, England : 1990)* 44(7), pp. 946-953.

Olinski, R. *et al.* 1992. DNA base modifications in chromatin of human cancerous tissues. *FEBS Letters* 309(2), pp. 193-198.

Oliveras-Ferraros, C. *et al.* 2011. Stem cell property epithelial-to-mesenchymal transition is a core transcriptional network for predicting cetuximab (Erbitux™) efficacy in KRAS wild-type tumor cells. *J Cell Biochem* 112(1), pp. 10-29.

Omura, K. 2003. Clinical implications of dihydropyrimidine dehydrogenase (DPD) activity in 5-FU-based chemotherapy: mutations in the &lt;i&gt;DPD&lt;/i&gt; gene, and DPD inhibitory fluoropyrimidines. *International Journal of Clinical Oncology* 8(3), pp. 132-138.

O'Neil, B. H. and McLeod, H. L. 2006. Thymidine Phosphorylase and Capecitabine: A Predictive Marker for Therapy Selection? *Journal of Clinical Oncology* 24(25), pp. 4051-4053.

Osborne, C. 1998. Steroid hormone receptors in breast cancer management. *Breast Cancer Res Treat* 51, pp. 227 - 238.

Paillas, S. *et al.* 2011. Targeting the p38 MAPK Pathway Inhibits Irinotecan Resistance in Colon Adenocarcinoma. *Cancer Research* 71(3), pp. 1041-1049.

Papp, J. *et al.* 2007. Germline MLH1 and MSH2 mutational spectrum including frequent large genomic aberrations in Hungarian hereditary non-polyposis colorectal cancer families: implications for genetic testing. *World J Gastroenterology* 13(19), pp. 2727-2732.

Pare, L. *et al.* 2008. Pharmacogenetic prediction of clinical outcome in advanced colorectal cancer patients receiving oxaliplatin/5-fluorouracil as first-line chemotherapy. *Br J Cancer* 99(7), pp. 1050-1055.

Parker, A. R. *et al.* 2005. Cells with pathogenic biallelic mutations in the human MUTYH gene are defective in DNA damage binding and repair. *Carcinogenesis* 26(11), pp. 2010-2018.

Parl, F. *et al.* 1984. Prognostic significance of estrogen receptor status in breast cancer in relation to tumor stage, axillary node metastasis, and histopathologic grading. *Cancer* 54, pp. 2237 - 2242.

Parsons, J. L. *et al.* 2005. APE1-dependent repair of DNA single-strand breaks containing 3'-end 8-oxoguanine. *Nucleid Acids Res* 33(7), pp. 2204-2209.

Parsons, R. *et al.* 1993. Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* 75(6), pp. 1227-1236.

Partlin, M. M. *et al.* 2003. Interactions of the DNA mismatch repair proteins MLH1 and MSH2 with c-MYC and MAX. *Oncogene*, 22(6), pp. 819-825.

Patt, Y. Z. *et al.* 2007. Capecitabine Plus 3-Weekly Irinotecan (XELIRI Regimen) as First-Line Chemotherapy for Metastatic Colorectal Cancer: Phase II Trial Results. *American Journal of Clinical Oncology* 30(4), pp. 350-357.

Paunio, T. *et al.* 1996. Preimplantation diagnosis by whole-genome amplification, PCR amplification, and solid-phase minisequencing of blastomere DNA. *Clin Chem* 42(9), pp. 1382-1390.

Pearce, M. *et al.* 2009. Mutations profiling in tumor samples using the Sequenom OncoCarta Panel. *Nature Methods* 6, pp. 7-8.

Peck, J. W. *et al.* 2002. The RhoA-binding protein, rhophilin-2, regulates actin cytoskeleton organisation. *J Biol Chem* 277(46), pp. 43924-43932.

Peeters, M. *et al.* 2009. Association of progression-free survival, overall survival, and patient-reported outcomes by skin toxicity and KRAS status in patients receiving panitumumab monotherapy. *Cancer* 115(7), pp. 1544-1554.

Peeters, M. *et al.* 2010. Use of massively parallel, next-generation sequencing to identify gene mutations beyond KRAS that predict response to panitumumab in a randomized, phase 3, monotherapy study of metastatic colorectal cancer (mCRC). *Am Assoc Cancer Res*, abstr LB-174.

Pegoraro, R. *et al.* 1986. Estrogen and progesterone receptors in breast cancer among women of different racial groups. *Cancer Res* 46(4 Pt 2), pp. 2117 - 2120.

Peltomäki, P. 2005. Lynch Syndrome Genes. *Familial Cancer* 4(3), pp. 227-232.

Peng, D. *et al.* 1996. Anti-Epidermal Growth Factor Receptor Monoclonal Antibody 225 Up-Regulates p27KIP1 and Induces G1 Arrest in Prostatic Cancer Cell Line DU145. *Cancer Research* 56(16), pp. 3666-3669.

Penney, K. L. *et al.* 2010. Genome-wide association study of prostate cancer mortality. *Cancer Epidemiol Biomarkers Prev*, 19(11), pp. 2869-2876.

Perrone, F. *et al.* 2009. PI3KCA/PTEN deregulation contributes to impaired responses to cetuximab in metastatic colorectal cancer patients. *Ann Oncol* 20(1), pp. 84-90.

Perrotte, P. *et al.* 1999. Anti-epidermal Growth Factor Receptor Antibody C225 Inhibits Angiogenesis in Human Transitional Cell Carcinoma Growing Orthotopically in Nude Mice. *Clinical Cancer Research* 5(2), pp. 257-264.

Personeni, N. *et al.* 2008. Clinical Usefulness of EGFR Gene Copy Number as a Predictive Marker in Colorectal Cancer Patients Treated with Cetuximab: A Fluorescent In situ Hybridization Study. *Clinical Cancer Research* 14(18), pp. 5869-5876.

Petit, A.M. *et al.* 1997. Neutralizing Antibodies against Epidermal Growth Factor and ErbB-2lneu Receptor Tyrosine Kinases Down-Regulate Vascular Endothelial Growth Factor Production by Tumor Cells in Vitro and in Vivo; Angiogenic Implications for Signal Transduction Solid Tumors. *Am J Pathol* 151(6), pp. 1523-1530.

Pichon, M. *et al.* 1996. Prognostic value of steroid receptors after long-term follow-up of 2257 operable breast cancers. *Br J Cancer* 73, pp. 1545 - 1551.

Pino, M. S. and Chung, D. C. 2010. The Chromosomal Instability Pathway in Colon Cancer. *Gastroenterology* 138(6), pp. 2059-2072.

Pinz, K. G. and Bogenhagen, D. F. 2006. The influence of the DNA polymerase γ accessory subunit on base excision repair by the catalytic subunit. *DNA Repair (Amst)*, 5(1), pp. 121-128.

Pittman, A. M. *et al.* 2009. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Research* 19(6), pp. 987-993.

Pittman, A. M. *et al.* 2009. The CDH1-160C>A polymorphism is a risk factor for colorectal cancer. *Int J Cancer* 125(7), pp. 1622-1625.

Pittman, A. M. *et al.* 2010. Allelic Variation at the 8q23.3 Colorectal Cancer Risk Locus Functions as a Cis-Acting Regulator of EIF3H. *PLoS Genet* 6(9), pp. e1001126.

Plotz, G. *et al.* 2006. DNA mismatch repair and Lynch syndrome. *Journal of Molecular Histology* 37(5), pp. 271-283.

Poehlmann, A. *et al.* 2007. K-ras mutation detection in colorectal cancer using the Pyrosequencing technique. *Pathology - Research and Practice* 203(7), pp. 489-497.

Polakis, P. 2000. Wnt signaling and cancer. *Genes & Development* 14(15), pp. 1837-1851.

Pomerantz, M. M. *et al.* 2009. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* 41(8), pp. 882-884.

Popat, S. and Houlston, R. S. 2005. A systematic review and meta-analysis of the relationship between chromosome 18q genotype, DCC status and colorectal cancer prognosis. *European journal of cancer (Oxford, England : 1990)* 41(14), pp. 2060-2070.

Popat, S. *et al.* 2005. Systematic Review of Microsatellite Instability and Colorectal Cancer Prognosis. *Journal of Clinical Oncology* 23(3), pp. 609-618.

Popat, S. *et al.* 2000. Prevalence of the APC E1317Q variant in colorectal cancer patients. *Cancer letters* 149(1-2), pp. 203-206.

Potter, J. D. 1999. Colorectal Cancer: Molecules and Populations. *Journal of the National Cancer Institute* 91(11), pp. 916-932.

Poulogiannis, G. *et al.* 2010. DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology* 56(2), pp. 167-179.

Powell, S. M. *et al.* 1992. APC mutations occur early during colorectal tumorigenesis. *Nature* 359(6392), pp. 235-237.

Poynter, J. N. *et al.* 2007. Variants on 9p24 and 8q24 Are Associated with Risk of Colorectal Cancer: Results from the Colon Cancer Family Registry. *Cancer Research* 67(23), pp. 11128-11132.

Pratilas, C. and Solit, D. 2007. Therapeutic strategies for targeting BRAF in human cancer. *Rev Recent Clin Trials* 2(2), pp. 121 - 134.

Prenen, H. *et al.* 2009. PIK3CA Mutations Are Not a Major Determinant of Resistance to the Epidermal Growth Factor Receptor Inhibitor Cetuximab in Metastatic Colorectal Cancer. *Clin Cancer Res* 15(9), pp. 3184-3188.

Preto, A. et al. 2008. BRAF provides proliferation and survival signals in MSI colorectal carcinoma cells displaying BRAFV600E but not KRAS mutations. *The Journal of Pathology* 214(3), pp. 320-327.

Pritchard, C. et al. 2010. COLD-PCR enhanced melting curve analysis improves diagnostic accuracy for KRAS mutations in colorectal carcinoma. *BMC Clinical Pathology* 10(1), p. 6.

Pritchard, C. C. and Grady, W. M. 2011. Colorectal cancer molecular biology moves into clinical practice. *Gut* 60(1), pp. 116-129.

Pueyo, G. et al. 2010. Cetuximab May Inhibit Tumor Growth and Angiogenesis Induced by Ionizing Radiation: A Preclinical Rationale for Maintenance Treatment After Radiotherapy. *Oncologist* 15(9), pp. 976-986.

Purcell, S. et al. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81(3), pp. 559-575.

Rachet, B. et al. 2009. Population-based cancer survival trends in England and Wales up to 2007: an assessment of the NHS cancer plan for England. *Lancet Oncol* 10(4), pp. 351-369.

Radicella, J. P. et al. 1997. Cloning and characterization of hOGG1, a human homolog of the OGG1 gene of Saccharomyces cerevisiae. *Proc Natl Acad Sci USA*, 94(15), pp. 8010-8015.

Rajagopalan, H. et al. 2003. The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer* 3(9), pp. 695-701.

Rakitina, T. V. et al. 2003. Additive Interaction of Oxaliplatin and 17-Allylamino-17-demethoxygeldanamycin in Colon Cancer Cell Lines Results from Inhibition of Nuclear Factor κB Signaling. *Cancer Research* 63(24), pp. 8600-8605.

Rass, U. et al. 2007. Actions of Aprataxin in Multiple DNA Repair Pathways. J Biol Chem, 282(13), pp. 9469-9474.

Razis, E. et al. 2008. Potential value of PTEN in predicting cetuximab response in colorectal cancer: An exploratory study. BMC Cancer 13(8), pp. 8234.

Rea, D. W. et al. 2005. A phase I/II and pharmacokinetic study of irinotecan in combination with capecitabine as first-line therapy for advanced colorectal cancer. Annals of Oncology 16(7), pp. 1123-1132.

Reeves, G. K. et al. 2007. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ 335(7630), p. 1134.

Roldán-Arjona, T. et al. 1997. Molecular cloning and functional expression of a human cDNA encoding the antimutator enzyme 8-hydroxyguanine-DNA☐glycosylase. Proceedings of the National Academy of Sciences 94(15), pp. 8016-8020.

Rosa, M. D. et al. 2003. The mutation spectrum of the APC gene in FAP patients from southern Italy: Detection of known and four novel mutations. Human Mutation 21(6), pp. 655-656.

Rosenquist, T. A. et al. 2003. The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. DNA Repair (Amst.), 2(5), pp. 581-591.

Rosenwald, I. B. et al. 1993. Elevated levels of cyclin D1 protein in response to increased expression of eukaryotic initiation factor eIF4E. Mol Cell Biol 13(12), pp. 7358-7363.

Rosenwald, I. B. et al. 1999. Upregulation of protein synthesis initiation factor eIF-4E is an early event during colon carcinogenesis. Oncogene 18(15), pp. 2507-2517.

Roth, A. D. *et al.* 2009. Correlation of molecular markers in colon cancer with stage-specific prognosis: Results of the translational study on the PETACC 3 - EORTC 40993-SAKK 60-00 trial. *Proc Am Soc Clin Oncol*, abstr 288.

Roth, A. D. *et al.* 2010. Prognostic role of *KRAS* and *BRAF* in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J Clin Oncol* 28(3), pp. 466-474.

Rothenberg, M. L. 1997. Topoisomerase I inhibitors: Review and update. *Annals of Oncology* 8(9), pp. 837-855.

Russo, A. *et al.* 2005. The TP53 Colorectal Cancer International Collaborative Study on the Prognostic and Predictive Significance of p53 Mutation: Influence of Tumor Site, Type of Mutation, and Adjuvant Treatment. *Journal of Clinical Oncology* 23(30), pp. 7518-7528.

Ruzzo, A. *et al.* 2007. Pharmacogenetic Profiling in Patients With Advanced Colorectal Cancer Treated With First-Line FOLFOX-4 Chemotherapy. *Journal of Clinical Oncology* 25(10), pp. 1247-1254.

Sachs, M. S. and Geballe, A. P. 2006. Downstream control of upstream open reading frames.*Genes Dev* 20(8), pp. 915-921.

Saif, M. W. *et al.* 2010. Safety and Efficacy of Panitumumab Therapy After Progression With Cetuximab: Experience at Two Institutions. *Clinical Colorectal Cancer* 9(5), pp. 315-318.

Sakumi, K. *et al.* 1993. Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP, a mutagenic substrate for DNA synthesis. *Journal of Biological Chemistry* 268(31), pp. 23524-23530.

Salomon, D. S. *et al.* 1995. Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit Rev Onco Hematol* 19(3), pp. 183-232.

Saltz, L. B. *et al.* 2000. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *N Engl J Med* 343(13), pp. 905-914.

Saltz, L. B. *et al.* 2003. The presence and intensity of the cetuximab-induced acne-like rash predicts increased survival in studies across multiple malignancies. *Proc Am Soc Clin Oncol* 22, abstr 817.

Saltz, L. B. *et al.* 2008. Bevacizumab in Combination With Oxaliplatin-Based Chemotherapy As First-Line Therapy in Metastatic Colorectal Cancer: A Randomized Phase III Study. *Journal of Clinical Oncology* 26(12), pp. 2013-2019.

Saltz, L. B. *et al.* 2007. Randomized Phase II Trial of Cetuximab, Bevacizumab, and Irinotecan Compared With Cetuximab and Bevacizumab Alone in Irinotecan-Refractory Colorectal Cancer: The BOND-2 Study. *Journal of Clinical Oncology* 25(29), pp. 4557-4561.

Saltz, L. B. *et al.* 2004. Phase II Trial of Cetuximab in Patients With Refractory Colorectal Cancer That Expresses the Epidermal Growth Factor Receptor. *Journal of Clinical Oncology* 22(7), pp. 1201-1208.

Samowitz, W. *et al.* 2005. Poor Survival Associated with the BRAF V600E Mutation in Microsatellite-Stable Colon Cancers. *Cancer Res* 65(14), pp. 6063 - 6069.

Samowitz, W. S. *et al.* 2000. Relationship of Ki-ras Mutations in Colon Cancers to Tumor Location, Stage, and Survival: A Population-based Study. *Cancer Epidemiology Biomarkers & Prevention* 9(11), pp. 1193-1197.

Sampson, J. R. *et al.* 2003. Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *Lancet* 362(9377), pp. 39-41.

Samuels, Y. and Ericson, K. 2006. Oncogenic PI3K and its role in cancer. *Current Opinion in Oncology* 18(1), pp. 77-82.

Samuels, Y. *et al.* 2004. High Frequency of Mutations of the PIK3CA Gene in Human Cancers. *Science* 304(5670), p. 554.

Sanger, F. *et al.* 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74(12), pp. 5463-5467.

Santini, D. *et al.* 2008. High Concordance of KRAS Status Between Primary Colorectal Tumors and Related Metastatic Sites: Implications for Clinical Practice. *Oncologist*, pp. theoncologist.2008-0181.

Santoro, I. M. and Groden, J. 1997. Alternative Splicing of the APC Gene and Its Association with Terminal Differentiation. *Cancer Research* 57(3), pp. 488-494.

Santucci-Darmamin, S. *et al.* 2002. The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. *Human Molecular Genetics*, 11(15), pp. 1697-1706.

Sartore-Bianchi, A. *et al.* 2009. PIK3CA Mutations in Colorectal Cancer Are Associated with Clinical Resistance to EGFR-Targeted Monoclonal Antibodies. *Cancer Res* 69(5), pp. 1851 - 1857.

Sartore-Bianchi, A. *et al.* 2009. Multi-Determinants Analysis of Molecular Alterations for Predicting Clinical Benefit to EGFR-Targeted Monoclonal Antibodies in Colorectal Cancer. *PLoS One* 4(10), e7287.

Sathiakumar, N. *et al.* 1998. Using the National Death Index to obtain underlying cause of death codes. *J Occup Environ Med* 40, pp. 808 - 813.

Scaltriti, M. and Baselga, J. 2006. The epidermal growth factor receptor pathway: A model for targeted therapy. *Clin Cancer Res* 12(18), pp. 5268-5272.

241

Scartozzi, M. *et al.* 2007. Epidermal growth factor receptor (EGFR) downstream signalling pathway in primary colorectal tumours and related metastatic sites: optimising EGFR-targeted treatment options. *Br J Cancer* 97(1), pp. 92-97.

Scheithauer, W. *et al.* 2003. Oral capecitabine as an alternative to i.v. 5-fluorouracil-based adjuvant therapy for colon cancer: safety results of a randomized, phase III trial. *Annals of Oncology* 14(12), pp. 1735-1743.

Schellens, J. H. M. 2007. Capecitabine. *Oncologist* 12(2), pp. 152-155.

Schreiber, V. *et al.* 2002. Poly(ADP-ribose) Polymerase-2 (PARP-2) Is Required for Efficient Base Excision DNA Repair in Association with PARP-1 and XRCC1. *J Biol Chem*, 277(25), pp. 23028-23036.

Schubbert, S. *et al.* 2007. Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer* 7(7).

Sclabas, G. M. *et al.* 2003. Restoring Apoptosis in Pancreatic Cancer Cells by Targeting the Nuclear Factor-[kappa]B Signaling Pathway With the Anti-Epidermal Growth Factor Antibody IMC-C225. *Journal of Gastrointestinal Surgery* 7(1), pp. 37-43.

Seiwert, T. *et al.* 2009. The MET receptor tyrosine kinase is a potential novel therapeutic target for head and neck squamous cell carcinoma. *Cancer Res* 69(7), pp. 3021 - 3031.

Seki, N. *et al.* 2002. Expression of eukaryotic initiation factor 4E in atypical adenomatous hyperplasia and adenocarcinoma of the human peripheral lung. *Clin Cancer Res* 8(10), pp. 3046-3053.

Shaheen, R. M. *et al.* 2001. Inhibited growth of colon cancer carcinomatosis by antibodies to vascular endothelial and epidermal growth factor receptors. *Br J Cancer* 85(4), pp. 584-589.

Shibutani, S. *et al.* 1991. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* 349(6308), pp. 431-434.

Shih, I.-M. *et al.* 2001. Evidence That Genetic Instability Occurs at an Early Stage of Colorectal Tumorigenesis. *Cancer Research* 61(3), pp. 818-822.

Shirota, Y. *et al.* 2001. ERCC1 and Thymidylate Synthase mRNA Levels Predict Survival for Colorectal Cancer Patients Receiving Combination Oxaliplatin and Fluorouracil Chemotherapy. *Journal of Clinical Oncology* 19(23), pp. 4298-4304.

Sidorenko, V. S. *et al.* 2009. Substrate specificity and excision kinetics of natural polymorphic variants and phosphomimetic mutants of human 8-oxoguanine-DNA glycosylase. *FEBS Journal* 276(18), pp. 5149-5162.

Sieber, O. M. *et al.* 2002. Whole-gene APC deletions cause classical familial adenomatous polyposis, but not attenuated polyposis or "multiple" colorectal adenomas. *Proceedings of the National Academy of Sciences* 99(5), pp. 2954-2958.

Sieber, O. M. *et al.* 2003. Multiple Colorectal Adenomas, Classic Adenomatous Polyposis, and Germ-Line Mutations in MYH. *New England Journal of Medicine* 348(9), pp. 791-799.

Siena, S. *et al.* 2009. Randomized phase III study of panitumumab with FOLFOX4 compared to FOLFOX4 alone as 1st-line treatment (tx) for metastatic colorectal cancer (mCRC): the PRIME trial. *Proc Am Soc Clin Oncol*, abstr 283.

Slattery, M. L. *et al.* 2010. Increased Risk of Colon Cancer Associated with a Genetic Polymorphism of SMAD7. *Cancer Research* 70(4), pp. 1479-1485.

Slupska, M. *et al.* 1996. Cloning and sequencing a human homolog (hMYH) of the Escherichia coli mutY gene whose function is required for the repair of oxidative DNA damage. *J. Bacteriol.* 178(13), pp. 3885-3892.

Smith, R. and Good, B. 2003. Chemoprevention of breast cancer and the trials of the National Surgical Adjuvant Breast and Bowel Project and others. *Endocr Relat Cancer* 10, pp. 347 - 357.

Sobol, R. W. *et al.* 1996. Requirement of mammalian DNA polymerase-β in base-excision repair. *Nature*, 379(6561), pp. 183-186.

Sobrero, A. F. *et al.* 2008. EPIC: Phase III Trial of Cetuximab Plus Irinotecan After Fluoropyrimidine and Oxaliplatin Failure in Patients With Metastatic Colorectal Cancer. *Journal of Clinical Oncology* 26(14), pp. 2311-2319.

Solit, D. B. *et al.* 2006. BRAF mutation predicts sensitivity to MEK inhibition. *Nature* 439(7074), pp. 358-362.

Søreide, K. *et al.* 2009. Microsatellite instability and DNA ploidy in colorectal cancer. *Cancer* 115(2), pp. 271-282.

Sotelo, J. *et al.* 2010. Long-range enhancers on 8q24 regulate c-Myc. *Proceedings of the National Academy of Sciences* 107(7), pp. 3001-3005.

Sparks, A. B. *et al.* 1998. Mutational Analysis of the APC/β-Catenin/Tcf Pathway in Colorectal Cancer. *Cancer Research* 58(6), pp. 1130-1134.

Steenken, S. and Jovanovic, S. V. 1997. How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *Journal of the American Chemical Society* 119(3), pp. 617-618.

Stemmler, M. P. *et al.* 2005. E-cadherin intron 2 contains cis-regulatory elements essential for gene expression. *Development*, 132(5), pp. 965-976.

Stoehlmacher, J. *et al.* 2004. A multivariate analysis of genomic polymorphisms: prediction of clinical outcome to 5-FU//oxaliplatin combination chemotherapy in refractory colorectal cancer. *Br J Cancer* 91(2), pp. 344-354.

Stoler, D. L. *et al.* 1999. The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proceedings of the National Academy of Sciences* 96(26), pp. 15121-15126.

Storey, A. *et al.* 1998. Role of a p53 polymorphism in the development of human papilloma-virus-associated cancer. *Nature* 393(6682), pp. 229-234.

Sunada, H. *et al.* 1986. Monoclonal antibody against epidermal growth factor receptor is internalized without stimulating receptor phosphorylation. *Proceedings of the National Academy of Sciences* 83(11), pp. 3825-3829.

Takahashi, K. and Suzuki, K. 1994. DNA synthesis-associated nuclear exclusion of p53 in normal human breast epithelial cells in culture. *Oncogene* 9(1), pp. 183-188.

Takeichi, M. 1991. Cadherin cell adhesion receptors as a morphogenetic regulator. *Science* 251(5000), pp. 1451-1455.

Tamburini, J. *et al.* 2009. Protein synthesis is resistant to rapamycin and constitutes a promising. *Blood* 114(8), pp. 1618-1627.

Tanaka, H. *et al.* 2006. BRAF mutation, CpG island methylator phenotype and microsatellite instability occur more frequently and concordantly in mucinous than non-mucinous colorectal cancer. *Int J Cancer* 118(11), pp. 2765 - 2771.

Tassinari, D. *et al.* 2010. EGFR inhibitors, bevacizumab, and chemotherapy in the treatment of metastatic colorectal cancer. *Proc Am Soc Clin Oncol,* abstr 436.

Tejpar, S. *et al.* 2010. Prognostic and predictive biomarkers in resected colon cancer: current status and future perspectives for integrating genomics into biomarker discovery. *Oncologist* 15(4), pp. 390-404.

Tenesa, A. and Dunlop, M. G. 2009. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat Rev Genet* 10(6), pp. 353-358.

245

Tenesa, A. *et al.* 2006. Association of MUTYH and colorectal cancer. *Br J Cancer*, 95(2), pp. 239-242.

Tenesa, A. *et al.* 2010. Ten common genetic variants associated with colorectal cancer risk are not associated with survival after diagnosis. *Clin Cancer Res* 16(14), pp. 3754-3759.

Tenesa, A. *et al.* 2008. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 40(5), pp. 631-637.

Terdiman, J. P. 2000. Genomic events in the adenoma to carcinoma sequence. *Semin Gastrointest Dis* 11(4), pp. 194-206.

Tetsu, O. and McCormick, F. 1999. β-Catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* 398(6726), pp. 422-426.

Thibodeau, S. *et al.* 1993. Microsatellite instability in cancer of the proximal colon. *Science* 260(5109), pp. 816-819.

Thomas, R. *et al.* 2007. High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 39(3), pp. 347 - 351.

Timmermann, B. *et al.* 2010. Somatic Mutation Profiles of MSI and MSS Colorectal Cancer Identified by Whole Exome Next Generation Sequencing and Bioinformatics Analysis. *PLoS ONE* 5(12), p. e15661.

Tol, J. *et al.* 2009. Chemotherapy, Bevacizumab, and Cetuximab in Metastatic Colorectal Cancer. *New England Journal of Medicine* 360(6), pp. 563-572.

Tomkinson, A. E. *et al.* 2001. Completion of base excision repair by mammalian DNA ligases. *Prog Nucleic Acid Res Mol Biol* 68, pp. 151-164.

Tomlinson, I. *et al.* 2007. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39(8), pp. 984-988.

Tomlinson, I. P. M. *et al.* 2008. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 40(5), pp. 623-630.

Tonra, J. R. *et al.* 2006. Synergistic Antitumor Effects of Combined Epidermal Growth Factor Receptor and Vascular Endothelial Growth Factor Receptor-2 Targeted Therapy. *Clinical Cancer Research* 12(7), pp. 2197-2207.

Topisirovic, I. *et al.* 2004. Phosphorylation of the Eukaryotic Translation Initiation Factor eIF4E Contributes to Its Transformation and mRNA Transport Activities. *Cancer Res* 64, pp. 8639-8642.

Tost, J. *et al.* 2003. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* 35(1), pp. 152-156.

Toyota, M. *et al.* 1999. CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences* 96(15), pp. 8681-8686.

Tsanou, E. *et al.* 2008. The E-cadherin adhesion molecule and colorectal cancer. A global literature approach. *Anticancer Res*, 28(6A), pp. 3815-3826.

Tsiatis, A. C. *et al.* 2010. Comparison of Sanger Sequencing, Pyrosequencing, and Melting Curve Analysis for the Detection of KRAS Mutations: Diagnostic and Clinical Implications. *The Journal of Molecular Diagnostics* 12(4), pp. 425-432.

Tsuzuki, T. *et al.* 2001. Spontaneous tumorigenesis in mice defective in the MTH1 gene encoding 8-oxo-dGTPase. *Proc Natl Acad Sci USA*, 98(20), pp. 11456-11461.

Tsuzuki, T. *et al.* 2001b. Analysis of MTH1 gene function in mice with targeted mutagenesis. *Mutat Res*, 477(1-2), pp. 71-78.

Tuupanen, S. *et al.* 2008. Allelic Imbalance at rs6983267 Suggests Selection of the Risk Allele in Somatic Colorectal Tumor Evolution. *Cancer Research* 68(1), pp. 14-17.

Tuupanen, S. *et al.* 2009. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 41(8), pp. 885-890.

Twelves, C. 2002. Capecitabine as first-line treatment in colorectal cancer: pooled data from two large, phase III trials. *European journal of cancer (Oxford, England : 1990)* 38, pp. 15-20.

Untawale, S. *et al.* 1993. Transforming growth factor-α production and autoinduction in a colorectal carcinoma cell line (DiFi) with an amplified epidermal growth factor receptor gene. *Cancer Res* 53, pp. 1630-1636.

Van Cutsem, E. *et al.* 2000. Capecitabine, an Oral Fluoropyrimidine Carbamate With Substantial Activity in Advanced Colorectal Cancer: Results of a Randomized Phase II Study. *Journal of Clinical Oncology* 18(6), pp. 1337-1345.

Van Cutsem, E. *et al.* 2007. Open-Label Phase III Trial of Panitumumab Plus Best Supportive Care Compared With Best Supportive Care Alone in Patients With Chemotherapy-Refractory Metastatic Colorectal Cancer. *Journal of Clinical Oncology* 25(13), pp. 1658-1664.

Van Cutsem, E. *et al.* 2008. An open-label, single-arm study assessing safety and efficacy of panitumumab in patients with metastatic colorectal cancer refractory to standard chemotherapy. *Annals of Oncology* 19(1), pp. 92-98.

Van Cutsem, E. *et al.* 2008. KRAS status and efficacy in the first-line treatment of patients with metastatic colorectal cancer (mCRC) treated with FOLFIRI with or without cetuximab: The CRYSTAL experience. *J Clin Oncol (Meeting abstracts)* 26, pp. 2.

Van Cutsem, E. *et al.* 2005. Oral capecitabine: Bridging the Atlantic divide in colon cancer treatment. *Seminars in oncology* 32(1), pp. 43-51.

Van Puijenbroek, M. *et al.* 2005. Mass spectrometry-based loss of heterozygosity analysis of single-nucleotide polymorphism loci in paraffin embedded tumors using the MassEXTEND assay: single-nucleotide polymorphism loss of heterozygosity analysis of the protein tyrosine phosphatase receptor type J in familial colorectal cancer. *J Mol Diagn* 7(5), pp. 623 - 630.

Viguier, J. *et al.* 2005. ERCC1 Codon 118 Polymorphism Is a Predictive Factor for the Tumor Response to Oxaliplatin/5-Fluorouracil Combination Chemotherapy in Patients with Advanced Colorectal Cancer. *Clinical Cancer Research* 11(17), pp. 6212-6217.

Visel, A. *et al.* 2009. Genomic views of distant-acting enhancers. *Nature* 461(7261), pp. 199-205.

Vivante, A. *et al.* 2007. High-throughput, sensitive and quantitative assay for the detection of BCR-ABL kinase domain mutations. *Leukemia* 21(6), pp. 1318 - 1321.

Vo, A. T. *et al.* 2005. hMRE11 deficiency leads to microsatellite instability and defective DNA mismatch repair. *EMBO Rep*, 6(5), pp. 438-444.

Vodenicharov, M. D. *et al.* 2000. Base excision repair is efficient in cells lacking poly(ADP-ribose) polymerase 1. *Nucleic Acids Res*, 28(20), pp. 3887-3896.

Vogelstein, B. and Kinzler, K. W. 1999. Digital PCR. *Proceedings of the National Academy of Sciences* 96(16), pp. 9236-9241.

Vogelstein, B. and Kinzler, K. W. 2002. The genetics basis of human cancer. McGraw Hill Higher Education.

Vogelstein, B. and Kinzler, K. W. 2004. Cancer genes and the pathways they control. *Nat Med* 10(8), pp. 789-99.

Walko, C. M. and Lindley, C. 2005. Capecitabine: A review. *Clinical Therapeutics* 27(1), pp. 23-44.

Wallis, Y. and Macdonald, F. 1996. The genetics of inherited colon cancer. *Clinical Molecular Pathology* 49(2), pp. M65-M73.

Walther, A. *et al.* 2008. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut* 57(7), pp. 941-950.

Walther, A. *et al.* 2009. Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer* 9(7), pp. 489-499.

Wang, C. *et al.* 2003. Prognostic Significance of Microsatellite Instability and Ki-ras Mutation Type in Stage II Colorectal Cancer. *Oncology* 64(3), pp. 259-265.

Wang, Z. Q. *et al.* 1997. PARP is important for genomic stability but dispensable in apoptosis. *Genes Dev*, 11(18), pp. 2347-2358.

Weichert, W. *et al.* 2010. KRAS Genotyping of Paraffin-Embedded Colorectal Cancer Tissue in Routine Diagnostics: Comparison of Methods and Impact of Histology. *The Journal of Molecular Diagnostics* 12(1), pp. 35-42.

Wheeler, J. M. *et al.* 2001. Hypermethylation of the promoter region of the E-cadherin gene (CDH1) in sporadic and ulcerative colitis associated colorectal cancer. *Gut* 48(3), pp. 367-371.

Wiederhold, L. *et al.* 2004. AP Endonuclease-Independent DNA Base Excision Repair in Human Cells. *Mol Cell*, 15(2), pp. 209-220.

Wilke, H. *et al.* 2008. Cetuximab Plus Irinotecan in Heavily Pretreated Metastatic Colorectal Cancer Progressing on Irinotecan: MABEL Study. *Journal of Clinical Oncology* 26(33), pp. 5335-5343.

Wilson Iii, D. M. and Bohr, V. A. 2007. The mechanics of base excision repair, and its relationship to aging and disease. *DNA Repair* 6(4), pp. 544-559.

Wolpin, B. M. and Mayer, R. J. 2008. Systemic Treatment of Colorectal Cancer. *Gastroenterology* 134(5), pp. 1296-1310.e1291.

Wong, E. *et al.* 2002. Mbd4 inactivation increases C→T transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci USA*, 99(23), pp. 14937-14942.

Wood, L. *et al.* 2007. The genomic landscapes of human breast and colorectal cancers. *Science* 318(5853), pp. 1108 - 1113.

Wood, R. D. *et al.* 2005. Human DNA Repair Genes. *Mutat Res*, 577(1-2), pp. 275-283.

Woodburn, J. R. The Epidermal Growth Factor Receptor and Its Inhibition in Cancer Therapy. *Pharmacology & Therapeutics* 82(2-3), pp. 241-250.

World Cancer Research Fund/American Institute for Cancer Research. (2007). Food, nutrition, physical activity and the prevention of cancer: a global perspective. *American Institute for Cancer Research.*

Woynarowski, J. M. *et al.* 1998. Sequence- and Region-Specificity of Oxaliplatin Adducts in Naked and Cellular DNA. *Molecular Pharmacology* 54(5), pp. 770-777.

Wright, J. B. *et al.* 2010. Upregulation of c-MYC in Cis Through a Large Chromatin Loop Linked to a Cancer Risk-Associated SNP in Colorectal Cancer Cells. *Mol. Cell. Biol.*, pp. MCB.01384-01309.

Wu, X. *et al.* 1995. Apoptosis induced by an anti-epidermal growth factor receptor monoclonal antibody in a human colorectal carcinoma cell line and its delay by insulin. *The Journal of Clinical Investigation* 95(4), pp. 1897-1905.

Wu, X. *et al.* 1996. Involvement of p27KIP1 in G1 arrest mediated by an anti-epidermal growth factor receptor monoclonal antibody. *Oncogene* 12(7), pp. 1397-1403.

Wu, Y. *et al.* 2001. A role for MLH3 in hereditary nonpolyposis colorectal cancer. *Nat Genet* 29(2), pp. 137-138.

Xiao, R. and Boehnke, M. 2009. Quantifying and correcting for the winner's curse in genetic association studies. *Genet Epidemiol* 33(5), pp. 453-462.

Xie, Y. *et al.* 1999. Characterization of the Repeat-Tract Instability and Mutator Phenotypes Conferred by a Tn3 Insertion in RFC1, the Large Subunit of the Yeast Clamp Loader. *Genetics*, 151(2), pp. 499-509.

Xing, J. *et al.* 2011. GWAS-identified colorectal cancer susceptibility locus associates with disease progression. *Eur J Cancer* In Press, Corrected Proof.

Yakushiji, H. *et al.* 1997. Biochemical and physicochemical characterization of normal and variant forms of human MTH1 protein with antimutagenic activity. *Mutat Res*, 384(3), pp. 181-194.

Yan, J. *et al.* 1998. Ras Isoforms Vary in Their Ability to Activate Raf-1 and Phosphoinositide 3-Kinase. *J Biol Chem* 273(37), pp. 24052-24056.

Yang, A. D. *et al.* 2006. Epithelial to mesenchymal transition is induced by oxaliplatin-resistance in colorectal cancer cells. *Proc Am Soc Oncol.* Abstr 289.

Yarden, Y. and Sliwkowski, M. X. 2001. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* 2(2), pp. 127-137.

Yuan, F. *et al.* 2004. Evidence for involvement of HMGB1 protein in human DNA mismatch repair. *J Biol Chem*, 279(20), pp. 20935-20940.

Yuen, S. *et al.* 2002. Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia. *Cancer Res* 62(22), pp. 6451 - 6455.

Zanke, B. W. *et al.* 2007. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 39(8), pp. 989-994.

Zhang, L. *et al.* 2007. Individual Overexpression of Five Subunits of Human Translation Initiation Factor eIF3 Promotes Malignant Transformation of Immortal Fibroblast Cells. *J Biol Chem* 282(8), pp. 5790-5800.

Zhang, L. *et al.* 2008. An Oncogenic Role for the Phosphorylated h-Subunit of Human Translation Initiation Factor eIF3. *J Biol Chem* 283(35), pp. 24047-24060.

Zhang, Y. *et al.* 2008. Identification of Regulatory Factor X as a Novel Mismatch Repair Stimulatory Factor. *J Biol Chem*, 283(19), pp. 12730-12735.

Zimmer, S.G. *et al.* 2000. Translational control of malignancy: the mRNA cap-binding protein, eIF-4E, as a central regulator of tumor formation, growth, invasion and metastasis. *Anticancer Res* 20(3A), pp. 1343-1351.