Searching for genetic variants associated with survival in patients with metastatic colorectal cancer

A thesis submitted in candidature for the degree of

Doctor of Philosophy (PhD)

Matthew G. Summers

2019

Division of Cancer and Genetics

School of Medicine

Cardiff University

Supervisors:

Prof. Jeremy Cheadle

Prof. Valentina Escott-Price

Abstract

Background

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and fourth highest cause of cancer-related death worldwide. Although the five-year survival rate for patients with early stage CRC is high ($\sim$92%), it is considerably decreased ($\sim$12%) for metastatic CRC (mCRC). A number of somatic prognostic factors are established for CRC, but few germline prognostic biomarkers have been identified.

Materials and methods

Survival analyses were stratified by somatic mutation and MSI status, and GWAS analyses for germline variants were performed on data from 2671 patients from the COIN and COIN-B clinical trials, followed by *in silico* functional analyses of four germline variants significantly associated with survival and validation analyses for three of these variants.

Results

Mutations in *KRAS* (HR 1.45, 95% CI 1.30-1.61, P=1.9x10^{-11}), *BRAF* (HR 2.31, 95% CI 1.85-2.87, P=7.8x10^{-14}), *NRAS* (HR 1.44, 95% CI 1.09-1.90, P=0.01) and MSI-positive tumours (HR 1.86, 95% CI 1.22-2.83, P=4.0x10^{-3}) were found to confer poor prognosis. Two SNPs (rs9356458 at 6q27 and rs17560791 at 2q35) of genome-wide significance ($P<$ 5.0x10^{-8}) and two (rs241477 at 14q31.3 and rs4074683 at 7p11.2) suggestive of association ($P<$ 1.0x10^{-5}) with survival were identified. These variants did not replicate in validation analyses, possibly due to overinflated effect sizes caused by winner's curse in the discovery set, or through clinicopathological differences between the study and validation cohorts.

Conclusion

The work presented in this thesis has identified somatic mutations in four oncogenes and tumour MSI status as being significantly associated with survival in mCRC patients. Four novel germline variants were also identified as potential candidate prognostic biomarkers for mCRC. Further work is required using larger validation cohorts of stage-matched patients with similar clinical and prognostic covariates to determine whether these variants can be validated as robust prognostic biomarkers for mCRC.

Acknowledgements

I would like to thank the following;

My supervisors, Prof. Jeremy Cheadle and Prof. Valentina Escott-Price for their guidance and support.

Prof. Richard Houlston and Dr. Philip Law for their input and advice as co-authors of the published work resulting from this thesis, and their help and hospitality during my time at the Institute for Cancer Research.

Prof. Tim Maughan, Prof. Rick Kaplan and Dr. Chris Smith for their input and advice as co-authors of the published work resulting from this thesis.

David Fisher at the Medical Research Council for his help and advice regarding the survival analyses, and providing additional COIN and COIN-B clinical molecular patient data.

Dr. Amanda Phipps, Dr. Barbara Banbury, Tabitha Harrison and Yi Lin at the Fred Hutchinson Cancer Research Centre and Dr. Claire Palles at the Institute of Cancer and Genomic Sciences, University of Birmingham for their assistance in collecting validation cohort data.

Dr. Elena Meuser, Alex Georgiades, Dr. Marc Naven, Dr. Daniel Nelmes, Dr. Thomas Cushion, Dr. Zoe Hudson, Dr. Magda Meissner, Dr. Gareth Marlow, Dr. Martin McClarty, Dr. Hywel Williams, Dr. Matthew Mort, Dr. Hannah West, Dr. Laura Thomas, Dr. Sara Seifan, Dr. Ellie Rad, Victoria Gray, Katie Watts and Christopher Wills for all of their support.

My family, for their unwavering belief in me and their continued love and support.

Finally, the patients of the COIN and COIN-B trials for giving consent for their samples to be analysed, without whom this research would not have been possible.

Abbreviations

Table 1: Abbreviations for amino acids.

Full name	3 letter code	1 letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table 2: Other abbreviations.

Abbreviation	Full name
%	Per cent
5-FU	5-fluorouracil
95% CI	95% confidence interval
A	Adenine
ACF	Aberrant crypt focus
AJCC	American Joint Committee on Cancer
ALKP	Alkaline phosphatase
ANOVA	Analysis of variance
APC	*Adenomatous polyposis coli*
ARCCA	Advanced Research Computing at Cardiff
BAT-25	Mononucleotide repeat marker
BAT-26	Mononucleotide repeat marker
BRAF	*v-Raf murine sarcoma viral oncogene homolog B*
C	Cytosine
c.	Coding
CAPOX	Chemotherapy regimen (oxaliplatin, capecitabine)
CCT6A	*Chaperonin Containing TCP1 Subunit 6A*
CD/CV	Common disease common variant
CDH1	*Cadherin-1*
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
COIN	COntinuous versus INtermittent
CRAN	Comprehensive R Archive Network
CRC	Colorectal cancer
CTU	Clinical Trials Unit
DNA	Deoxyribonucleic acid
DTD	Diagnosis to death
EGFR	*Epidermal Growth Factor Receptor*
ELOVL5	*ELOVL Fatty Acid Elongase 5*
eQTL	Expression quantitative trait loci
FAP	Familial adenomatous polyposis
FDR	False discovery rate
FFPE	Formalin fixed paraffin embedded
FKBP9P1	*FKBP Prolyl Isomerase 9 Pseudogene 1*
FOLFIRI	Chemotherapy regimen (folinic acid, fluorouracil, irinotecan)
FOLFOX	Chemotherapy regimen (folinic acid, fluorouracil, oxaliplatin)
FP	Fluoropyrimidine

FPRP	False positive report probability
G	Guanine
GBAS	*Protein NipSnap homolog 2*
GECCO	Genetics and Epidemiology of Colorectal Cancer Consortium
GLM	Generalised linear model
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study
HER2	*Human epidermal growth factor receptor 2*
HER3	*Human epidermal growth factor receptor 3*
HNPCC	Hereditary non-polyposis colon cancer
HPC	High-performance cluster
HPFS	Health Professionals Follow-up Study
HR	Hazard ratio
HWE	Hardy-Weinberg Equilibrium
IBD	Identity by descent
ICR	Institute of Cancer Research
IDE	integrated development environment
IGFBP2	*Insulin Like Growth Factor Binding Protein 2*
IGFBP5	*Insulin Like Growth Factor Binding Protein 5*
KRAS	*Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*
LANCL2	*LanC Like 2*
LD	Linkage disequilibrium
MAF	Minor allele frequency
MAP	MUTYH-associated polyposis
MAPK	Mitogen-activated protein kinase
MARCH4	*E3 ubiquitin-protein ligase MARCH4*
mCRC	Metastatic colorectal cancer
MET	*MET Proto-Oncogene, Receptor Tyrosine Kinase*
MIR1913	*MicroRNA 1913*
MLH1	*MutL homolog 1*
MMR	Mismatch repair
moAB	monoclonal antibody
MPC1	*Mitochondrial pyruvate carrier 1*
MRC	Medical Research Council
MREC	Medical Research and Ethics Committee
mRNA	Messenger RNA
MRPS17	*Mitochondrial Ribosomal Protein S17*
MSH3	*MutS Homolog 3*
MSI	Microsatellite instability
MSS	Microsatellite stable

NA	Not available
NAT	Natural antisense transcript
NCBI	National Centre for Biotechnology Information
NF-kB	Nuclear Factor Kappa B Subunit 1
NGS	Next Generation Sequencing
NHS	Nurses' Health Study
NRAS	*Neuroblastoma RAS viral oncogene homolog*
NUPR2	*Nuclear Protein 2 Transcriptional Regulator*
OS	Overall survival
OxMdG	Oxaliplatin modified de Gramont
p	Short arm of chromosome (from French petite)
P	P-value
p.	Protein
PD	Progressive disease
PD-1	Programmed Cell Death Protein 1
PD-L1	Programmed Cell Death Ligand 1
PFS	Progression-free survival
PHS	Physician's Health Study
P^{HET}	P-value of heterogeneity
PHKG1	*Phosphorylase Kinase Catalytic Subunit Gamma 1*
PI3K	*Phosphoinositide 3-kinase*
PIK3CA	*Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha*
POLE	*DNA polymerase epsilon catalytic subunit*
PRR18	*Proline Rich 18*
PS	Performance status
PSPH	*Phosphoserine Phosphatase*
PSPHP1	*Phosphoserine Phosphatase Pseudogene 1*
PTEN	*Phosphatase and tensin homologue*
q	Long arm of chromosome
Q-Q	Quantile-quantile
QC	Quality control
QUASAR 2	Quick and Simple and Reliable Trial 2
RECIST	Response Evaluation Criteria In Solid Tumours
REMARK	REporting recommendations for tumour MARKer prognostic studies
RPL37A	*Ribosomal Protein L37a*
RPS6KA2	*Ribosomal Protein S6 Kinase A2*
SCOT	Short Course Oncology Therapy
SD	Standard deviation
SEPT14	*Septin 14*
SFT2D1	*SFT2 Domain Containing 1*

SMAD4	*Mothers against decapentaplegic homolog 4*
SMARCAL1	*SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin Subfamily A Like 1*
SNORA15	*Small Nucleolar RNA, H/ ACA Box 15*
SNP	Single nucleotide polymorphism
SUMF2	*Sulfatase Modifying Factor 2*
T	Thymine
TF	Transcription factor
TGFα	*Transforming growth factor alpha*
TNP1	*Transition protein 1*
TP53	*Tumor protein P53*
TSG	Tumour suppressor gene
TUBBP6	*Tubulin Beta Class I Pseudogene 6*
UCSC	University of California Santa Cruz
UICC	Union for International Cancer Control
VEGF	*Vascular endothelial growth factor*
VICTOR	Vioxx in Colorectal Cancer Therapy: Definition of Optimal Regimen
VITAL	VITamins And Lifestyle Study
VOPP1	*VOPP1 WW Domain Binding Protein*
WBC	White blood cell
WHI	Women's Health Initiative
WHO	World Health Organization
XELOX	Chemotherapy regimen (oxaliplatin, capecitabine)
XRCC5	*X-ray repair cross-complementing protein 5*
ZNF713	*Zinc Finger Protein 713*

List of Figures

List of Tables

Publications

Publications as direct outcome of the work presented in this thesis:

Matthew G. Summers, Christopher G. Smith, Timothy S. Maughan, Richard Kaplan, Valentina Escott-Price and Jeremy P. Cheadle. *BRAF* and *NRAS* Locus-Specific Variants Have Different Outcomes on Survival to Colorectal Cancer. *Clinical Cancer Research* (2017), 23(11); 2742-9. DOI: 10.1158/1078-0432.CCR-16-1541. PMID: 27815357.

Additional work I have been involved in during the course of this PhD project:

Victoria Gray, Sarah Briggs, Claire Palles, Emma Jaeger, Timothy Iveson, Rachel Kerr, Mark P Saunders, James Paul, Andrea Harkin, John McQueen, **Matthew G. Summers**, Elaine Johnstone, Haitao Wang, Laura Gatcombe, Timothy S. Maughan, Richard Kaplan, Valentina Escott-Price, Nada A. Al-Tassan, Brian F. Meyer, Salma M. Wakil, Richard S. Houlston, Jeremy P. Cheadle, Ian Tomlinson, David N. Church. Pattern Recognition Receptor Polymorphisms as Predictors of Oxaliplatin Benefit in Colorectal Cancer. *JNCI: Journal of the National Cancer Institute* (2019), 111(8):828-836. DOI: https://doi.org/10.1093/jnci/djy215. PMID: 30649440.

Matthew G. Summers, Timothy S. Maughan, Richard Kaplan, Philip J. Law, Richard S. Houlston, Valentina Escott-Price and Jeremy Cheadle. Comprehensive analysis of colorectal cancer risk loci and survival outcome, a prognostic role for *CDH1* variants. *European Journal of Cancer* (2019), 124:56-63. DOI: https://doi.org/10.1016/j.ejca.2019.09.024. PMID: 31734605.

Contents

Chapter 1

Introduction

1.1 Colorectal cancer

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and fourth highest cause of cancer-related death worldwide (Brenner et al., 2014). Over the course of the next decade, the global CRC burden is expected to increase by 60% to more than 2.2 million new cases and 1.1 million deaths (Arnold et al., 2017). This rise in CRC incidence and mortality has been attributed to a number of factors including the increasingly ageing population, unfavourable modern dietary habits and an increase in risk factors such as smoking, low physical exercise and obesity, particularly in developed countries (Kuipers et al., 2015).

In Europe, nearly 50% of CRC patients will develop metastases over the course of their disease and approximately 25% of patients present with metastases at initial diagnosis (Van Cutsem et al., 2014). This can be attributed to CRC often developing without symptoms until it has reached an advanced stage, which delays diagnosis and therefore negatively influences patient prognosis (Draht et al., 2018). This contributes to the high mortality rates reported for the disease; the current five-year survival rate for CRC is approximately 65% (Van Cutsem et al., 2014). In the UK, there are approximately 41,700 new cases of CRC diagnosed annually, which equates to approximately 114 new cases every day. There are approximately 16,000 CRC related deaths in the UK every year, an average of approximately 44 per day (Cancer Research UK, 2018).

In the advanced disease setting, the clinical outcome for Stage IV, or metastatic CRC (mCRC), patients has improved over the last decade. The median overall survival (OS) of patients with mCRC treated in phase III trials and large observational series or registries is 30 months; more than double that of 20 years ago (Van Cutsem et al., 2016). This improvement may largely be attributed to an increase in the number of patients being referred for and undergoing surgical resection, earlier detection of metastatic disease and a more strategic approach towards the delivery of systemic therapy (Van Cutsem et al., 2016).

However, the average five-year survival rate for CRC patients (65%) is still low in comparison to a number of other cancers including breast (91%), skin (92%) and prostate cancers (99%) (Miller et al., 2019). The prognosis for early stage CRC patients is generally good, with survival rates for Stage I and II patients being 91% and 81%, respectively. This is in stark contrast to mCRC patients, for which five-year survival is currently 12% (Miller et al., 2019).

It is clear that more research is required in order to help improve the prognosis of patients with mCRC. CRC is a heterogeneous disease (Hanahan and Weinberg, 2011), and the molecular mechanisms underlying CRC development are clinically important because they are related to the prognosis and treatment response of the patient (Brenner et al., 2014).

Genetic prognostic and predictive markers for CRC can be of both a somatic (acquired) (Karapetis et al., 2008; Walther et al., 2009) and germline (inherited) nature (Lynch and de la Chapelle, 2003; Jasperson et al., 2010). Currently, one of the primary goals of CRC research is to translate knowledge of genetic biomarkers into diagnostic tools that can aid in the clinical management of the disease (Bacolod and Barany, 2011). Novel genetic prognostic biomarkers may therefore lead to a more effective way of combating CRC (Bacolod and Barany, 2011) and it is imperative that further research into novel prognostic biomarkers is conducted. This may aid clinicians in treating patients in the most effective way and ultimately improve the prognosis of patients with mCRC.

1.1.1 Staging of CRC

The gold standard for classifying patients with cancer, defining prognosis and determining the best treatment approaches is the American Joint Committee on Cancer (AJCC) staging manual, currently in its 8th edition (Amin et al., 2017). Using this staging system, colorectal cancers are classified according to invasion depth (T stage), lymph node involvement (N stage) and presence of distant metastases (M stage) (Table 1.1).

Table 1.1: Tumour, node and metastasis (TNM) classification of colorectal cancer.

Local invasion depth (T stage)	
Tx	No information about local tumour infiltration available
Tis	Tumour restricted to mucosa, no infiltration of lamina muscularis mucosae (carcinoma in situ)
T1	Infiltration through lamina muscularis mucosae into submucosa, no infiltration of lamina muscularis propria
T2	Infiltration into, but not beyond, lamina muscularis propria
T3	Infiltration into subserosa or non-peritonealised pericolic or perirectal tissue, or both; no infiltration of serosa or neighbouring organs
T4a	Infiltration of the serosa
T4b	Infiltration of neighbouring tissues or organs
Lymph node involvement (N stage)	
Nx	No information about lymph node involvement available
N0	No lymph node involvement
N1a	Cancer cells detectable in one regional lymph node
N1b	Cancer cells detectable in two to three regional lymph nodes
N1c	Tumour satellites in subserosa or pericolic or perirectal fat tissue, regional lymph nodes not involved
N2a	Cancer cells detectable in four to six regional lymph nodes
N2b	Cancer cells detectable in seven or more regional lymph nodes
Presence of distant metastases (M stage)	
Mx	No information about distant metastases available
M0	No distant metastases detectable
M1a	Metastasis to one distant organ or distant lymph nodes
M1b	Metastasis to more than one distant organ or set of distant lymph nodes or peritoneal metastasis
M1c	Metastasis to the peritoneal surface

Adapted from Brenner et al., 2014.

The stages described in Table 1.1 are combined into an overall CRC stage definition (Table 1.2), upon which therapeutic decisions are based (Brenner et al., 2014).

Table 1.2: Pathological staging of colorectal carcinoma, AJCC 8th Edition.

Stage	T stage	N stage	M stage
Stage 0	Tis	N0	M0
Stage I	T1-T2	N0	M0
Stage II	T3-T4	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T4a	N0	M0
Stage IIC	T4b	N0	M0
Stage III	Any	N+	M0
Stage IIIA	T1-T2	N1/N1c	M0
	T1	N2a	M0
Stage IIIB	T3-T4a	N1/N1c	M0
	T2-T3	N2a	M0
	T1-T2	N2b	M0
Stage IIIC	T4a	N2a	M0
	T3-T4a	N2b	M0
	T4b	N1-N2	M0
Stage IV	Any	Any	M+
Stage IVA	any T	any N	M1a
Stage IVB	any T	any N	M1b
Stage IVC	any T	any N	M1c

Adapted from Brenner et al., 2014.

The primary role of the staging system used in the AJCC staging manual is seen by many as a standardised classification system of evaluating cancer at a population level in terms of the extent of disease (Amin et al., 2017). For decades, this staging system has been successfully deployed worldwide by the AJCC and its partner, the Union for International Cancer Control (UICC), and is the principal classifying system for patients with solid tumours (Kattan et al., 2016). This staging system is widely accepted due to its simplistic categorical nature, and the fact it is clinically useful to apply in patient management because it is associated with overall survival (Kattan et al., 2016). Stage IV CRC is also referred to as mCRC. The majority of patients with mCRC are beyond curative therapy, and instead are treated with palliative care (Van Cutsem et al., 2014).

1.2 Colorectal tumourigenesis

1.2.1 Causes of colorectal tumourigenesis

Unlike other commonly diagnosed cancers, no single risk factor accounts for the majority of CRC cases (Brenner et al., 2014). Colorectal tumourigenesis is caused by a number of contributory elements, including dietary and lifestyle factors (Kuipers et al., 2015) as well as genetic factors such as somatic and germline mutations (Lynch and de la Chapelle, 2003; Karapetis et al., 2008; Jasperson et al., 2010; Fearon, 2011). The most significant dietary and lifestyle risk factors for CRC have long been considered to be a diet consisting of high amounts of unsaturated fats and red meat, high total energy intake, excessive alcohol consumption and reduced physical activity (Potter, 1999; Slattery, 2000; Huxley et al., 2009). A family history of CRC (Lynch and de la Chapelle, 2003), inflammatory bowel disease (Jess et al., 2012) and cigarette smoking (Liang et al., 2009) have also been identified as risk factors for the development of CRC. Contrastingly, some factors have been shown to have a protective effect against CRC development, such as non-steroidal anti-inflammatory drugs, oestrogen and calcium (Hawk and Levin, 2005).

Significant progress has been made in the identification of molecular mechanisms underlying both inherited CRC syndromes and sporadic CRC development (Fearon, 2011). Inherited CRC syndromes account for $\sim$3–5% of all occurrences of CRC (Lynch and de la Chapelle, 2003; Jasperson et al., 2010). The most common familial syndromes predisposing to CRC development include familial adenomatous polyposis (FAP) (Fearnhead et al., 2001), MUTYH-associated polyposis (MAP) (Al-Tassan et al., 2002) and Lynch syndrome (Hereditary Non-Polyposis Colorectal Cancer, HNPCC) (Lynch and de la Chapelle, 2003). Despite strong hereditary components, the majority of CRCs develop sporadically over several years, sometimes decades via an accumulation of somatic alterations (Brenner et al., 2014).

1.2.2 Cancer genes

Two distinct types of genetic defect initiate colorectal tumourigenesis; alterations that lead to novel or increased function of oncogenes and alterations that lead to the loss of function of tumour-suppressor genes (TSGs) (Fearon, 2011). An oncogene is a gene which when mutated will become constitutively active (e.g. continually transcribed), or will cause the resultant mutant protein to become active when the wild type would not normally be. An activating somatic mutation in only one allele of an oncogene is generally all that is required in order for constitutive activation to occur (Knudson, 1997; Fearon, 2011). Oncogenes can become activated through a variety of causes including chromosomal translocations (caused by chromosomal instability, CIN), intragenic mutations of residues that are critical in regulating the activity of the resulting proteins, and gene amplifications (Vogelstein and Kinzler, 2004). Examples of oncogenes that are often mutated during colorectal tumourigenic processes are *KRAS* and *BRAF*, both of which are commonly mutated as part of the adenoma–carcinoma sequence (Fearon and Vogelstein, 1990; Fearon, 2011), described in Chapter 1.2.3.2.

TSG mutations contribute to tumourigenesis in the opposite way to those of oncogenes. The mutation of a TSG normally leads to reduced activity of the mutant protein. Causes of this inactivity include mutations that result in protein truncation, deletions or insertions (indels), missense mutations at residues that are essential for the activity of the protein, or epigenetic silencing (Vogelstein and Kinzler, 2004). An example of a TSG commonly inactivated in CRC is *Adenomatous Polyposis Coli* (*APC*). *APC* is possibly the most important TSG inactivated in CRC, as approximately 80% of sporadic CRCs contain biallelic *APC* mutations (Powell et al., 1992; Miyoshi et al., 1992; Fearnhead et al., 2001). The loss of APC function occurs early during the progression towards malignancy, and is one of the initiating steps of CRC (Groden et al., 1991; Powell et al., 1992).

Generally, mutations in both alleles of a TSG are required in order for inactivation to occur (i.e. they act in a recessive fashion) (Knudson, 1996; Fearon, 2011). In the case of sporadic CRC, these mutations are both somatic. This is in contrast to inherited CRC predisposition syndromes; for example in FAP, patients carry a germline mutation in *APC* and subsequently acquire a somatic mutation affecting the other *APC* allele, which initiates early-onset colorectal tumourigenesis (Fearnhead et al., 2001). This concurs with the two hit hypothesis proposed by Knudson (Knudson, 1996), as shown in Figure 1.1.

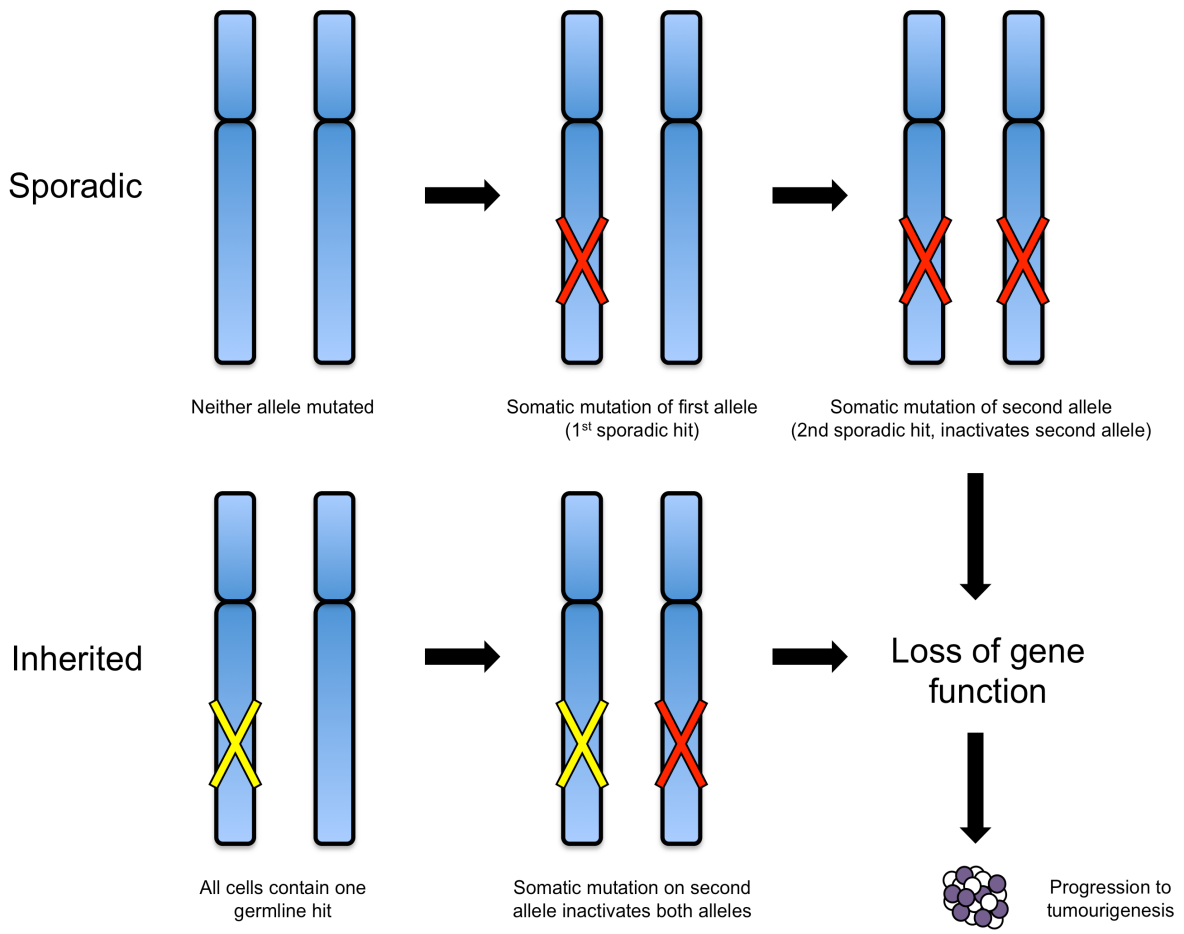

Figure 1.1: The two-hit hypothesis of tumourigenesis. In the case of sporadic CRC, two separate somatic alterations (either single nucleotide variations or loss of heterozygosity) are required to cause the loss of gene function that leads to tumourigenesis (one per allele). In the case of inherited CRC, every cell contains one germline mutation in the first allele; therefore only one somatic mutation is required for loss of gene function and progression to tumour formation (adapted from Knudson, 1996).

1.2.3 Genomic instability and pathways of colorectal tumourigenesis

1.2.3.1 Genomic instability

Genomic instability is a characteristic of almost all human cancers, resulting from mutations in DNA repair genes and driving cancer development (Negrini et al., 2010; Sansregret et al., 2018). Approximately 65–70% of CRCs present with CIN (Swanton et al., 2006; Walther et al., 2009; Turajlic et al., 2019), which describes the high frequency at which the chromosomes in cancer cells change in structure and number over time compared to normal cells (Negrini et al., 2010; Sansregret et al., 2018).

Abnormal structures and numbers of chromosomes and abnormal mitoses associated with CIN cause genetic alterations to occur, which can lead to tumourigenesis (Walther et al., 2008). CIN occurs due to a number of genetic changes that involve the activation of oncogenes (e.g. *KRAS*), and the inactivation of TSGs such as *APC* (Fearnhead et al., 2001; Swanton et al., 2006; Sansregret et al., 2018).

The second most common type of genomic instability is microsatellite instability (MSI). Microsatellites are short, repetitive DNA sequences found throughout the genome, which are prone to mutations such as indels and frameshifts (Sinicrope and Sargent, 2012). MSI is a characteristic of the hereditary disorder Lynch syndrome (Lynch and de la Chapelle, 2003), and is also found in approximately 15% of sporadic CRCs (Peltomaki, 2001; Haydon and Jass, 2002; Popat et al., 2005; Sinicrope and Sargent, 2012). MSI can develop as a result of mutations or epigenetic inactivation of DNA mismatch repair (MMR) genes (Peltomaki, 2001; Sinicrope and Sargent, 2012). MSI-positive tumours tend to occur mainly in the proximal colon (Popat et al., 2005).

A small number of CRCs show signs of genomic instability characterised by epigenetic silencing events involving neither CIN nor MSI, but a mechanism referred to as CpG Island Methylator Phenotype (CIMP) (Esteller, 2008; Dahlin et al., 2010). It should be noted that these three subtypes are not necessarily mutually exclusive and have been observed to overlap (Ogino et al., 2009).

1.2.3.2 The adenoma-carcinoma sequence

The adenoma-carcinoma sequence describes the process underlying CRC development (Fearon and Vogelstein, 1990; Fearon, 2011). The sequence begins with the formation of an adenoma from normal epithelial cells, normally through the acquisition of biallelic inactivating mutations in *APC*. Subsequent mutations in a combination of oncogenes and TSGs then cause the early adenoma to develop into an intermediate, then late adenoma, before eventually progressing to a carcinoma (Figure 1.2).

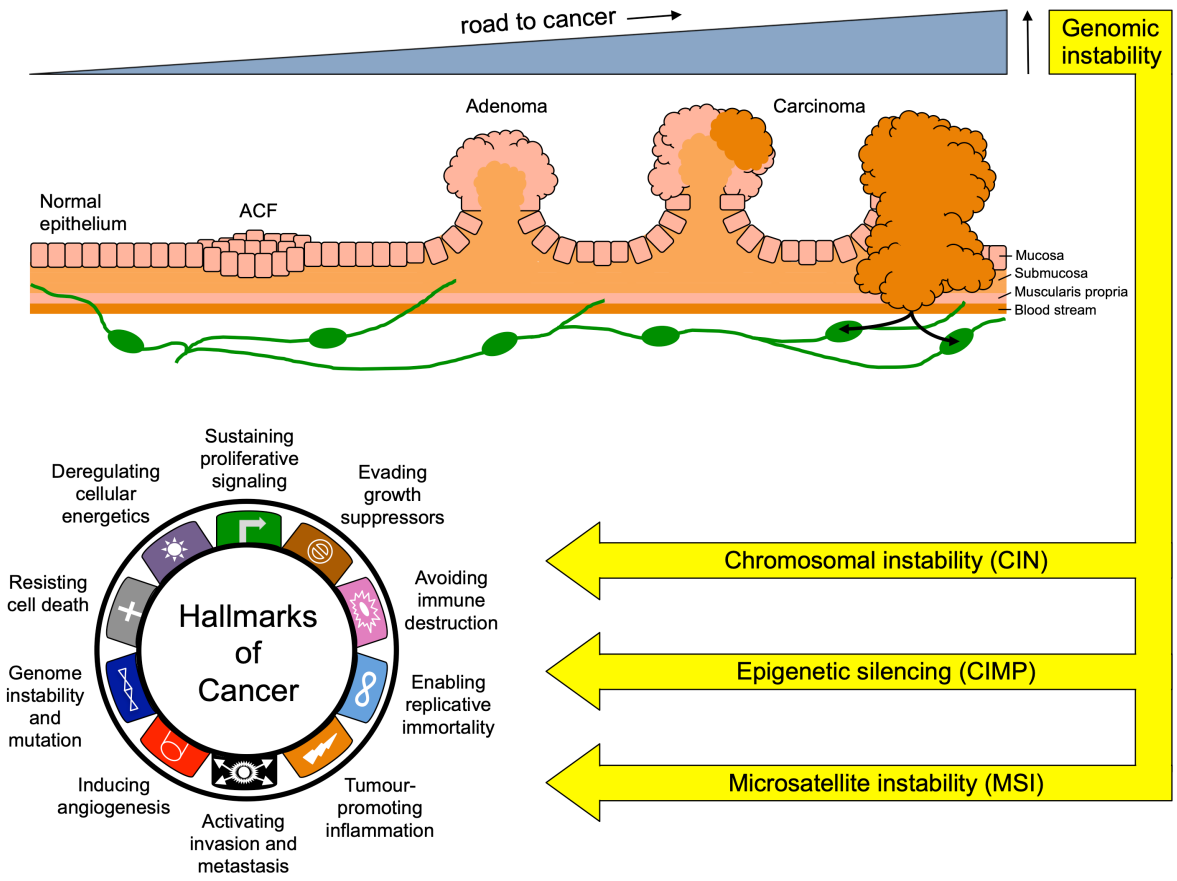

Figure 1.2: Colorectal tumourigenesis. Tumourigenesis follows a series of neoplastic changes resulting in the transformation of normal epithelium to aberrant crypt focus (ACF), early, late adenoma and invasive carcinoma (adapted from Walther et al., 2009 and Hanahan and Weinberg, 2011).

1.2.3.3 The *Adenomatous Polyposis Coli* gene

The *APC* gene encodes a large protein that plays an important role in the canonical Wnt signalling pathway. Germline *APC* mutations cause the dominant disease FAP (Fearnhead et al., 2001; Lynch and de la Chapelle, 2003), characterised by hundreds to thousands of adenomatous polyps in the large intestine and, if left untreated, inevitable development of CRC at a young age. Biallelic somatic mutations in *APC* are evident in approximately 80% of cases of sporadic CRC (Fearnhead et al., 2001). Mutations in *APC* occur early in CRC tumourigenesis (Powell et al., 1992) (Figure 1.2) and usually cause a truncation of the resulting protein, impacting its function as a negative regulator of canonical Wnt signalling.

1.2.3.4 The Epidermal Growth Factor Receptor (EGFR) signalling pathway

The EGFR signalling pathway is an important pathway that regulates cell growth and survival, and is central to human tumourigenesis; the function of EGFR is altered in a number of epithelial tumours, including CRC (Cardo-Vila et al., 2010; Bertotti et al., 2015). The EGFR signalling pathway is activated upon a ligand binding to the receptor. Although there are eight known ligands for EGFR, the majority of research into the signalling cascade post-ligand activation is focused on the effects of just two; epidermal growth factor (EGF) and transforming growth factor alpha (TGFα) (Henriksen et al., 2013). EGFR can also be activated by ligand-independent mechanisms, which can result in a variety of signalling outcomes. In the case of ligand-dependent EGFR activation, once a ligand has bound to the receptor, a number of signal transduction pathways downstream of the receptor are activated that initiate cellular proliferation or survival responses due to expression of EGFR target effectors, including the RAS-RAF-MEK-ERK cascade (otherwise known as the mitogen-activated protein kinase, or MAPK pathway) which includes KRAS, BRAF and NRAS, as well as phosphoinositide-3-kinase (PI3K) pathways (Huels et al., 2018) (Figure 1.3A).

Numerous mechanisms lead to the aberrant activation of EGFR signalling in cancer, including receptor mutation and overexpression as well as ligand-independent activation (Mendelsohn and Baselga, 2006). This aberrant behaviour plays a key role in the development and growth of tumour cells (Yarden, 2001). The EGFR signalling pathway is up-regulated in 60–80% of mCRCs (Goldstein and Armin, 2001), therefore anti-EGFR inhibiting targeted therapies are commonly used for the treatment of mCRC patients (Van Cutsem et al., 2014, 2016).

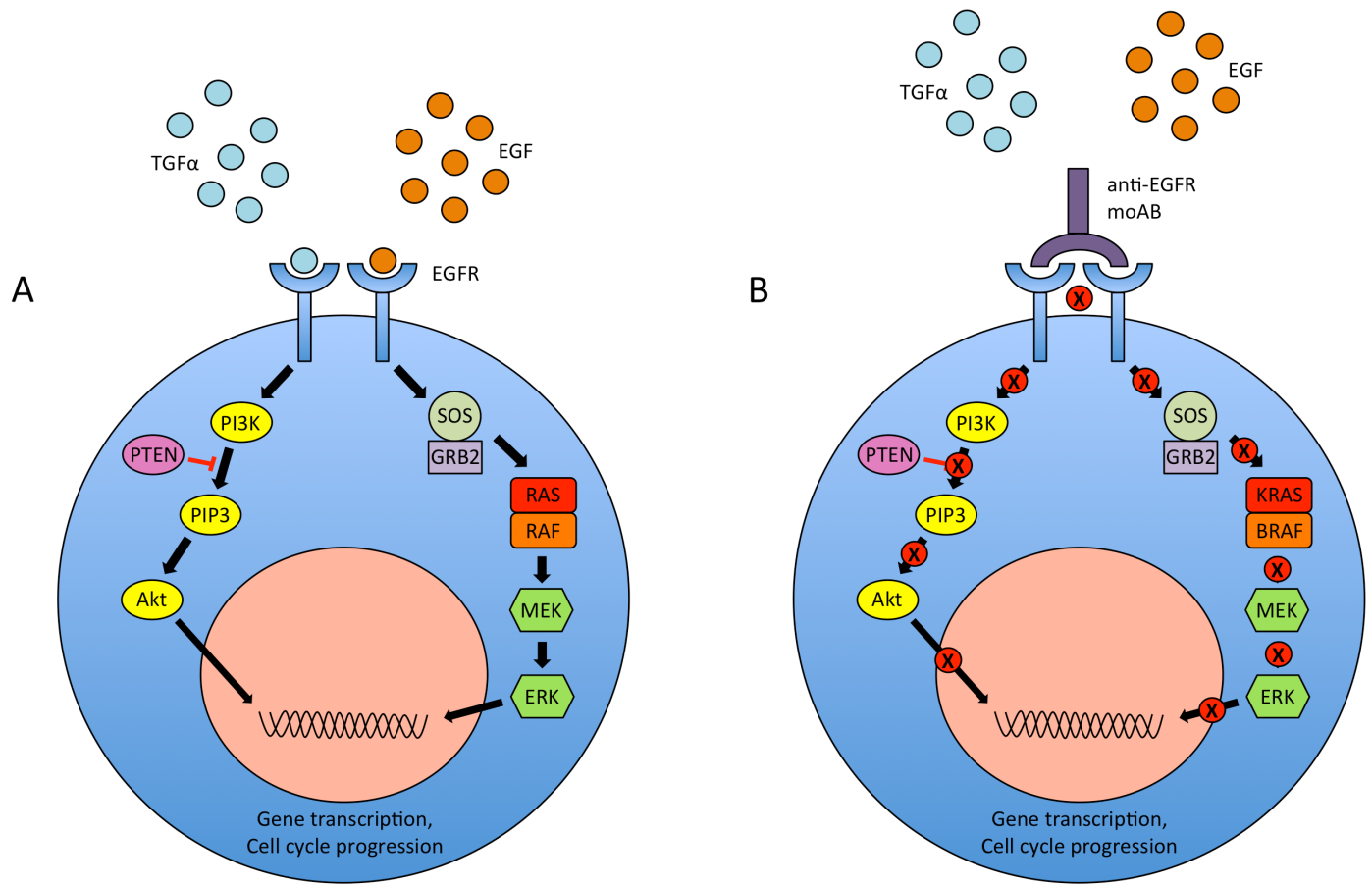

Figure 1.3: The EGFR signalling pathway. A: Ligands such as EGF and TGFα bind to the receptor, which triggers a cellular signalling cascade. This cascade contributes to the regulation of gene transcription and cell cycle progression. B: When a monoclonal antibody (moAB) has been administered (such as cetuximab), the moAB binds to the receptor and blocks ligand binding, thus inactivating the signalling cascade and preventing the tumour cell from aberrantly regulating gene transcription and cell cycle progression. Should a member of the pathway (for instance *KRAS* or *NRAS*) harbour an activating mutation, the signalling cascade activates from this point onwards, rendering the moAB ineffective. Therefore, patients are screened for *KRAS* and *NRAS* mutations, and cetuximab is administered to patients with *KRAS* and *NRAS* wild type tumours only (Van Cutsem et al., 2014, 2016).

1.3 Treatment of CRC

1.3.1 Cytotoxic agents

Clinical practice currently employs a combination of surgery and/or radiotherapy and chemotherapy in the treatment of CRC (Kuipers et al., 2015). In the case of mCRC, the backbone of first-line palliative chemotherapy (either administered alone or in combination with targeted agents) consists of a fluoropyrimidine (FP) either in the form of intravenous 5-fluorouracil (5-FU) or the oral FP capecitabine in various combinations and schedules (Douillard et al., 2000; de Gramont et al., 2000; Van Cutsem et al., 2014).

Combination chemotherapies comprised of this backbone include FOLFOX (or oxaliplatin modified de Gramont; OxMdG); a combination of 5-FU, folinic acid (leucovorin) and oxaliplatin administered intravenously, FOLFIRI; a combination of 5-FU, leucovorin and irinotecan administered intravenously; and CAPOX (or XELOX); a combination of intravenously administered oxaliplatin and the orally administered FP capecitabine (Cassidy et al., 2004; Madi et al., 2012; Van Cutsem et al., 2014). These drugs are cytotoxic (i.e. they are toxic to cells), and are therefore administered with the intention to damage and potentially kill rapidly dividing tumour cells. Unfortunately these therapies are not tumour-specific, therefore causing damage not only to cancer cells, but to healthy tissue as well. Side effects of these treatments can include severe nausea, vomiting, and diarrhoea as well as immunosuppression, hair loss, skin rashes and fatigue (Madi et al., 2012; Chionh et al., 2017).

1.3.2 Biological targeted agents

Biological targeted therapies are also used in combination with chemotherapy in order to improve the outcome of mCRC patients, which take into account the molecular genetic background of a patient's tumour and provide a significant step towards personalised treatment (Van Cutsem et al., 2014). MoABs target specific receptors in signalling pathways commonly dysregulated in mCRC; such as bevacizumab for the vascular endothelial growth factor (VEGF) signalling pathway (Tol et al., 2009; Simkens et al., 2015) and cetuximab for the EGFR signalling pathway (Maughan et al., 2011; Smith et al., 2013) (Figure 1.3A). In the case of EGFR, the moAB binds to the receptor, preventing ligands such as TGFα and EGF from binding, thus inhibiting downstream signalling and preventing hyperactive cell proliferation (Smith et al., 2013) (Figure 1.3B).

Side effects of these treatments can include hypomagnesaemia and allergic responses such as infusion reactions and acneiform rashes (Wolpin and Mayer, 2008). MoABs can improve both overall as well as progression-free survival (PFS), while also preserving the quality of life of CRC patients that do not respond to chemotherapy (Karapetis et al., 2008). Generally, a combination of cytotoxics and biological targeted treatments produce a higher median survival rate, although this comes at the cost of a more complicated management of treatment-related side effects than the administration of moAB monotherapy alone (Van Cutsem et al., 2014).

1.3.2.1 Somatic *RAS* mutations and the efficacy of anti-EGFR therapies

Despite the reduction of tumour proliferation achieved by moABs, some groups of mCRC patients do not benefit from these treatments due to somatic mutations in genes involved in the respective targeted signalling pathways (Amado et al., 2008; Karapetis et al., 2008; De Roock et al., 2010a). Examples of this are the *KRAS* and *NRAS* (*RAS*) genes, which are mutated in $\sim$40% and $\sim$4% of mCRC tumours, respectively (Smith et al., 2013). A somatic activating *RAS* mutation will cause the signalling cascade to activate downstream from this point, and thus uncontrolled cell proliferation can continue irrespective of whether a moAB has bound to the receptor, rendering the treatment ineffective (Amado et al., 2008; Karapetis et al., 2008; De Roock et al., 2010a). Consequently, cetuximab benefits only those patients with *RAS* wild type tumours, and therefore is only administered to this patient group in clinical practice (Amado et al., 2008; Karapetis et al., 2008; De Roock et al., 2010b; Van Cutsem et al., 2014, 2016).

1.3.3 Clinical trials in CRC

The efficacy of CRC treatment regimens are tested in clinical trials to determine whether a certain drug, or combination of drugs, performs better than another. There have been numerous clinical trials that tested the efficacy of standard chemotherapies and targeted agents for CRC. A number of these trials collected formalin fixed, paraffin embedded (FFPE) tumour and/or blood samples from consenting patients, which can be used for translational research into patient response to treatment and survival (Maughan et al., 2011; Smith et al., 2013). This project used patient samples from the COIN (COntinuous versus INtermittent) and COIN-B trials, which are described in detail in Chapter 2.1. A number of clinical trials assessing the efficacy of the most commonly administered CRC treatments are shown in Table 1.3.

Table 1.3: Clinical trials of commonly administered CRC treatments.

Trial name	N	Treatment	Tumour	Blood	Reference
ADD-ASPIRIN*	2600+	As	Yes	Yes	Coyle et al. 2016
CAIRO[m]	820	Ca, I, O	No	No	Koopman et al. 2007
CAIRO2[m]	529	Ca, O, B, Ce	Yes	No	Tol et al. 2009
CAIRO3[m]	558	Ca, B, O	No	No	Simkens et al. 2015
CAIRO4*[m]	360+	Fp, B	No	No	t Lam-Boer et al. 2014
CAIRO5*[m]	640+	Ff, Fx, B, Pa	Yes	No	Huiskens et al. 2015
COIN[m]	2445	O, Fp, Ce, Fl, Fo, Ca	Yes	Yes	Maughan et al. 2011
COIN-B[m]	226	Ce, Fx	Yes	Yes	Wasan et al. 2014
CORGI-L[m]	47	Ca, O	No	No	Gunnlaugsson et al. 2009
CRYSTAL[m]	1198	Ff	Yes	No	Van Cutsem et al. 2015
FOCUS[m]	2135	Fl, I, O	No	No	Seymour et al. 2007
FOCUS2[m]	459	O, Fl, Ca	No	No	Seymour et al. 2011
FOCUS3[m]	240	I, Fl, O, Ce, Be	Yes	No	Maughan et al. 2014
FOCUS4*[m]	4730+	As, Ca, Pa	Yes	No	Kaplan et al. 2013
FOXTROT	150	O, Fo, Fl, Pa	Yes	No	Foxtrot Collaborative Group 2012
MODUL*[m]	1400+	Fx, B,Ce, At, Ca, T, Pe, Co	Yes	Yes	Schmoll et al. 2018
MOSAIC	2246	Ox, Fl, Le	No	No	Andre et al. 2004
OPUS[m]	315	Ce, Fx	Yes	No	Bokemeyer et al. 2011
PICCOLO[m]	1198	Pa, I, Ci	Yes	No	Seymour et al. 2013
PRIME[m]	1183	Fx, Pa	Yes	No	Douillard et al. 2010
QUASAR	3239	Fl, Fo, L	No	No	Gray et al. 2007
QUASAR 2	4855	B, Ca	No	No	Rosmarin et al. 2014
SCOT	6088	Fx, Ca, O	No	No	Iveson et al. 2018
SOFT[m]	512	B, Fx, S	No	No	Nakamura et al. 2015
TRANSSCOT*	6144	Fx, Ca, O	Yes	Yes	Engelmann et al. 2016
VICTOR	2434	R	Yes	No	Midgley et al. 2010

N: Sample size. Tumour: FFPE tumour tissue collected. Blood: Blood DNA collected. As: Aspirin, At: Atezolizumab, B: Bevacizumab, Ca: Capecitabine, Ce: Cetuximab, Ci: Ciclosporin, Co: Cobimetinib, Ff: FOLFIRI, Fl: Fluorouracil, Fo: Folinic acid, Fp: Fluoropyrimidine, Fx: FOLFOX, I: Irinotecan, L: Levamisole, Le: Leucovorin, O: Oxaliplatin, Pa: Panitumamab, Pe: Petuzumab, R: Rofecoxib, S: SOX, T: Trastuzumab, V: Vemurafenib. All trials conducted in adjuvant disease unless otherwise stated. [m]: Trial conducted in metastatic disease. *: Trial on-going.+: Estimated sample size.

1.4 Factors influencing CRC prognosis

The prognosis of patients with CRC is highly dependent on the AJCC stage of the tumour at diagnosis (Walther et al., 2009). Patients who present with advanced disease generally have an inferior prognosis to those diagnosed at an earlier stage; the average five-year relative survival rate for mCRC (Stage IV) patients being 12%, compared to 91% and 82% for Stage I and II CRC patients, respectively (Miller et al., 2019). Relative survival also decreases with age, and at younger ages is slightly higher for women than for men (Brenner et al., 2014).

In terms of mCRC patients, other clinical parameters such as performance status (PS), alkaline phosphatase (ALKP) levels, white blood cell (WBC) count and number of metastatic sites have been reported to influence prognosis (Kohne et al., 2002). Due to their roles in numerous metastatic processes (Li, 2016), platelet counts are also intrinsically linked to CRC prognosis. Other factors that impact prognosis include increased central adiposity and a lack of regular physical activity prior to diagnosis (Haydon et al., 2006), systemic inflammatory response to the tumour (Leitch et al., 2007) and the tumour's immunologic environment (the type, density and location of immune cells within the tumour) (Galon et al., 2006).

A number of somatic factors have been shown to affect prognosis and response to treatment, and are used in clinical practice as prognostic and/or predictive biomarkers for the management of CRC patients (Labianca et al., 2013; Van Cutsem et al., 2014, 2016). *RAS* mutations have been shown to confer poor prognosis (Richman et al., 2009; Eklof et al., 2013; Schirripa et al., 2015) and are a negative predictive biomarker for therapeutic choices involving EGFR moAB therapies for mCRC (Amado et al., 2008; Karapetis et al., 2008; Douillard et al., 2013). *RAS* mutations are therefore recommended for testing as a predictive biomarker in the clinical management mCRC patients (Van Cutsem et al., 2014, 2016).

BRAF mutations confer a poor prognosis and are a prognostic marker in the clinical management of mCRC (Richman et al., 2009; Tran et al., 2011; Van Cutsem et al., 2011, 2016). Moreover, *BRAF* mutations are frequently associated with MSI (Venderbosch et al., 2014), which also was shown to confer a poor prognosis in mCRC patients (Tran et al., 2011; Maughan et al., 2011; Venderbosch et al., 2014). MSI is a predictive biomarker for the use of immune checkpoint inhibitors in the treatment of patients with mCRC (Van Cutsem et al., 2016). MSI is associated with good prognosis in earlier stages of CRC (Popat et al., 2005; Bertagnolli et al., 2009; Hutchins et al., 2011; Lochhead et al., 2013) and is used as a prognostic marker in the management of Stage II CRC patients (Labianca et al., 2013).

While a number of somatic biomarkers have been established, there are far fewer instances of validated germline biomarkers for CRC prognosis. This is potentially often due to small sample sizes, the selection of unsuitable candidates or the lack of an independent validation cohort to verify the findings of the original study (Smith et al., 2015; Phipps et al., 2016). The first commonly inherited germline variant robustly associated with survival in mCRC is the SNP rs9929218 (16q22.1), intronic to the *CDH1* gene, which encodes for E-cadherin.

Patients homozygous for the minor allele showed significantly poorer survival as compared to those who where heterozygous or homozygous for the major allele (HR 1.28, 95% CI 1.14-1.43, P=2.2x10^{-5}) through a candidate gene study (Smith et al., 2015). One further SNP of prognostic significance in CRC has also been validated, rs209489 (6p12.1), intronic to the *ELOVL5* gene, which encodes the fatty acid elongase ELOVL5. The minor allele of rs209489 was shown to be associated with a significantly poorer prognosis (HR 2.00, 95% CI 1.50-2.50, P=3.7x10^{-9}) through a genome-wide association study (GWAS) (Phipps et al., 2016).

Table 1.4: Factors shown to influence CRC prognosis.

Factor	Reference(s)
	Clinical factors
AJCC stage at diagnosis*	Amin et al. 2017; Miller et al. 2019
Age at diagnosis	Brenner et al. 2014
Sex	Majek et al. 2013; Brenner et al. 2014
ALKP levels	Kohne et al. 2002
WBC count	Kohne et al. 2002
Preoperative platelet count	Wan et al. 2013
Obstruction and perforation at presentation	Steinberg et al. 1986
Number of metastatic sites	Kohne et al. 2002
Resection status of primary tumour	Faron et al. 2015
WHO PS	Kemeny and Braun Jr. 1983; Kohne et al. 2002
	Somatic factors
KRAS mutations^{+}	Andreyev et al. 1998, 2001; Eklof et al. 2013
BRAF mutations*	Richman et al. 2009; Tran et al. 2011
NRAS mutations^{+}	Schirripa et al. 2015
MSI (early stage)*	Popat et al. 2005; Bertagnolli et al. 2009; Hutchins et al. 2011; Lochhead et al. 2013
MSI (late stage)	Tran et al. 2011; Smith et al. 2013
PIK3CA mutations	Ogino et al. 2009
TP53 mutations	Russo et al. 2005; Munro et al. 2005
Loss of heterozygosity at 18q	Ogunbiyi et al. 1998
SMAD4 protein and mRNA levels	Alhopuro et al. 2005
CIN	Walther et al. 2008
CIMP	Samowitz et al. 2005; Barault et al. 2008; Ogino et al. 2009

Germline factors	
rs9929218 genotype	Smith et al. 2015
rs209489 genotype	Phipps et al. 2016

AJCC: American joint Committee on Cancer. CIN: Chromosomal instability. MSI: Microsatellite instability. CIMP: CpG island methylator phenotype. ALKP: Alkaline phosphatase. WBC: White blood cell. WHO: World Health Organization. PS: Performance status. *: Currently in use as a prognostic factor in the clinical setting. +: Currently in use as a predictive marker for anti-EGFR therapy.

1.4.1 Emerging biomarkers

A number of biomarkers are emerging beyond *RAS* mutational status that may impact on the response to EGFR-targeted therapies (Van Cutsem et al., 2016). These include *HER2*, *MET* and *KRAS* gene amplification, ligands such as TGFα, amphiregulin and epiregulin, *EGFR* mutations and alterations/mutations in *HER3*, *PIK3CA* and *PTEN* (Van Cutsem et al., 2016). Recent studies have identified that DNA polymerase epsilon *POLE* proofreading domain mutations might define a subset of cancers, including CRC (Palles et al., 2013) and endometrial cancer (Church et al., 2015), that display a particularly favourable prognosis (Domingo et al., 2016). Somatic mutations of the *POLE* proofreading domain may therefore be a promising candidate biomarker for CRC (Domingo et al., 2016).

1.5 Genome-wide association studies (GWASs)

Until recently, the search for germline factors that affect CRC prognosis focused predominantly on candidate genes functioning within the pharmacological pathways of the chemotherapeutic agents used in the treatment of CRC (Smith et al., 2015). The limited robustness from prior studies may partly reflect the shortcomings of a candidate gene-based approach (the pathways, genes and SNPs most relevant to and most robustly associated with CRC prognosis may not have a previously understood role in CRC progression and survival) (Phipps et al., 2016). The advent of GWAS enabled researchers to look beyond the candidate gene approach and analyse the whole human genome in order to identify any heritable genetic variants associated with disease susceptibility and prognosis (Hirschhorn and Daly, 2005). The GWAS approach therefore represents an unbiased and comprehensive option that can be performed in the absence of convincing evidence regarding the location or function of the causal genes (Hirschhorn and Daly, 2005).

Genetic information obtained by genome mapping projects such as the International HapMap Project (Frazer et al., 2007) and the 1000 Genomes Project (Abecasis et al., 2012) have enabled GWAS analyses to determine associations between SNPs and traits that were previously impossible to identify. From a clinical standpoint, this information could serve to improve the

understanding of the biology of disease and thus potentially result in disease prevention or the development of more effective treatments (Visscher et al., 2017).

For over a decade, GWASs have been successful in finding germline variants associated with a phenotype of interest for a wide variety of diseases, including schizophrenia, type II diabetes and cystic fibrosis (Bush and Moore, 2012). Significant associations have also been identified for the development of a number of cancers, such as breast (Easton et al., 2007; Haiman et al., 2011), lung (Hu et al., 2011; Lan et al., 2012), gastric (Shi et al., 2011), prostate (Thomas et al., 2008) and colorectal cancers. To date, 79 SNPs have been associated with susceptibility to CRC through GWAS analyses (Law et al., 2019).

1.5.1 Underlying concepts of GWAS design

1.5.1.1 Single nucleotide polymorphisms (SNPs)

A SNP is a single base-pair change in the DNA sequence that has a minor allele frequency (MAF) greater than 1% in at least one population (Risch, 2000; Erichsen and Chanock, 2004). While the majority of SNPs are 'silent' and do not alter the expression of a gene (Erichsen and Chanock, 2004), in some instances SNPs can have an impact on gene expression, known as expression quantitative trait loci (eQTLs, described further in Chapter 1.6.1.2) (Abecasis et al., 2010). SNPs can also cause other functional consequences including amino acid changes, changes to mRNA transcript stability and transcription factor (TF) binding affinity (Griffith et al., 2008). SNPs are therefore used as markers of genomic regions in GWASs in order to ascertain the heritable quantitative traits that are risk factors for disease susceptibility and prognosis (Hirschhorn and Daly, 2005).

1.5.1.2 The 'common disease, common variant' (CD/CV) hypothesis

The CD/CV hypothesis states that common disorders are likely to be caused by genetic variants that are also common in the population (Bush and Moore, 2012). Following this hypothesis, a number of consequences emerge for the study of complex disease. Firstly, if common variants influence disease, the effect size for a single variant must be small relative to those found in rare disorders. This in turn means that the allele frequency and the prevalence of disease in the population are correlated, thus common variants cannot have large effect sizes (Bush and Moore, 2012). Secondly, if common alleles have modest effect sizes, but common disorders are heritable, it is suggested that common disease susceptibility results from the combined action of several common variants, meaning unrelated affected individuals share a significant amount of disease alleles (Wang et al., 2005). The allele frequencies of the variants analysed and their phenotypic effect sizes are therefore intrinsically linked to the potential statistical power of a GWAS, and thus ultimately the success of the study based on a pre-specified sample size (Wang et al., 2005).

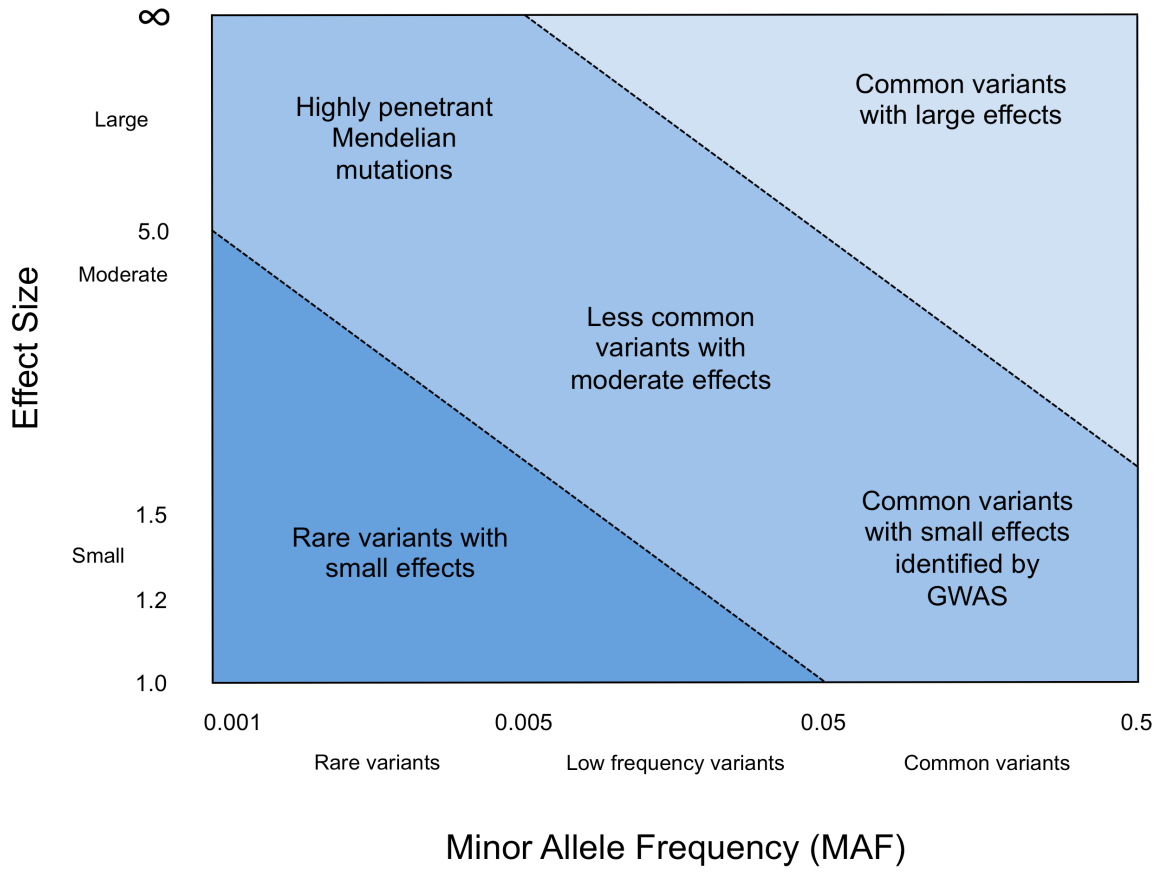

Figure 1.4: The spectrum of disease allele effects. The concept of disease association is often thought of in terms of MAF and effect size. Alleles with high penetrance for Mendelian disorders are extremely rare and have large effect sizes (top left corner), while the majority of findings from GWASs are associations of common SNPs with small effect sizes (bottom right corner). The diagonal lines represent the majority of discovered genetic associations (adapted from Bush and Moore, 2012).

1.5.1.3 Linkage disequilibrium (LD)

LD describes the non-random association between alleles at two different loci on the same chromosome (Slatkin, 2008) and is variable across the genome and different populations in a complex and unpredictable way (Hirschhorn and Daly, 2005). LD is based upon the concept of chromosomal linkage, where two markers on a chromosome remain physically joined through generations of recombination events (Bush and Moore, 2012). There are two measures of LD; D' and r^2. Generally, D' is used in population genetics, whereas r^2 is informative in association studies because it is inversely proportional to the sample size required for detecting disease association given a fixed genetic risk (Wang et al., 2005). The r^2 value is a statistical measure of correlation, scaled between 0 and 1. A high r^2 for two SNPs indicates that one allele of the first SNP is often observed with one allele of the second SNP; and as such the two SNPs are in

high LD with each other. In the case of high LD between two SNPs, only one of the two SNPs needs to be genotyped in order to capture the allelic variation of both. A high r^2 value can only be obtained if the alleles of two SNPs are correlated, occur on the same ancestral haplotype and have a similar MAF (Wang et al., 2005), which is why GWASs using common SNP arrays are often underpowered when attempting to detect associations of rare causal markers (Visscher et al., 2017). Utilising LD information prevents the genotyping of SNPs that would provide redundant information, and hence is often exploited to optimise GWAS analyses (Bush and Moore, 2012). Using data from the International HapMap project, it has been shown that a subset of between 500,000 to 1,000,000 SNPs across the genome should be sufficient to capture over 80% of commonly occurring SNPs in European populations (Li et al., 2008).

The presence of LD confers two potential outcomes regarding a SNP that has been found to be significantly associated with a phenotype. Either the SNP influencing the biological system that ultimately leads to the phenotype has been directly genotyped in the study and found to be statistically associated with the trait (known as direct association), or the influential SNP is not directly genotyped, but instead a tag SNP in high LD with the influential SNP is genotyped and statistically associated with the phenotype (known as indirect association) (Hirschhorn and Daly, 2005). Therefore, a SNP that has been found to be significantly associated with a trait through a GWAS should not automatically be assumed to be the causal SNP. Once an associated allele is discovered, a critical next step is to define causal allele through fine-mapping (Raychaudhuri, 2011). Following the CD/CV hypothesis, a panel of 500,000 to 1,000,000 SNPs will identify any common SNPs that are associated with common phenotypes. In order for this to be done efficiently and cost-effectively, state-of-the-art genotyping technology is required (Bush and Moore, 2012). Chip-based microarrays allow over a million SNPs to be tested for association, although the vast majority of these SNPs will not have been directly genotyped. The imputation of SNPs that have not been directly genotyped is a key underlying concept of GWAS study design that is a cost-effective way of producing the vast number of SNPs required to perform a successful GWAS.

1.5.2 Genotype imputation

Genotype imputation is the process of predicting genotypes that are not directly assayed in a sample of individuals. The imputation of unobserved SNPs can recover some of the genetic data lost due to imperfect LD between observed genotypes and causal variants (Visscher et al., 2017). This process uses a fully sequenced reference panel of haplotypes at a dense set of SNPs to impute genetic information into a study sample of individuals that have been genotyped at a subset of the SNPs. These *in silico* genotypes greatly increase the number of SNPs that can be tested for association. There are many benefits of genotype imputation in GWASs, such as increasing the statistical power of the study, enabling fine-mapping of the causal variant and facilitating meta-analysis of the data (Marchini and Howie, 2010). Figure 1.5 highlights the fundamental process of genotype imputation.

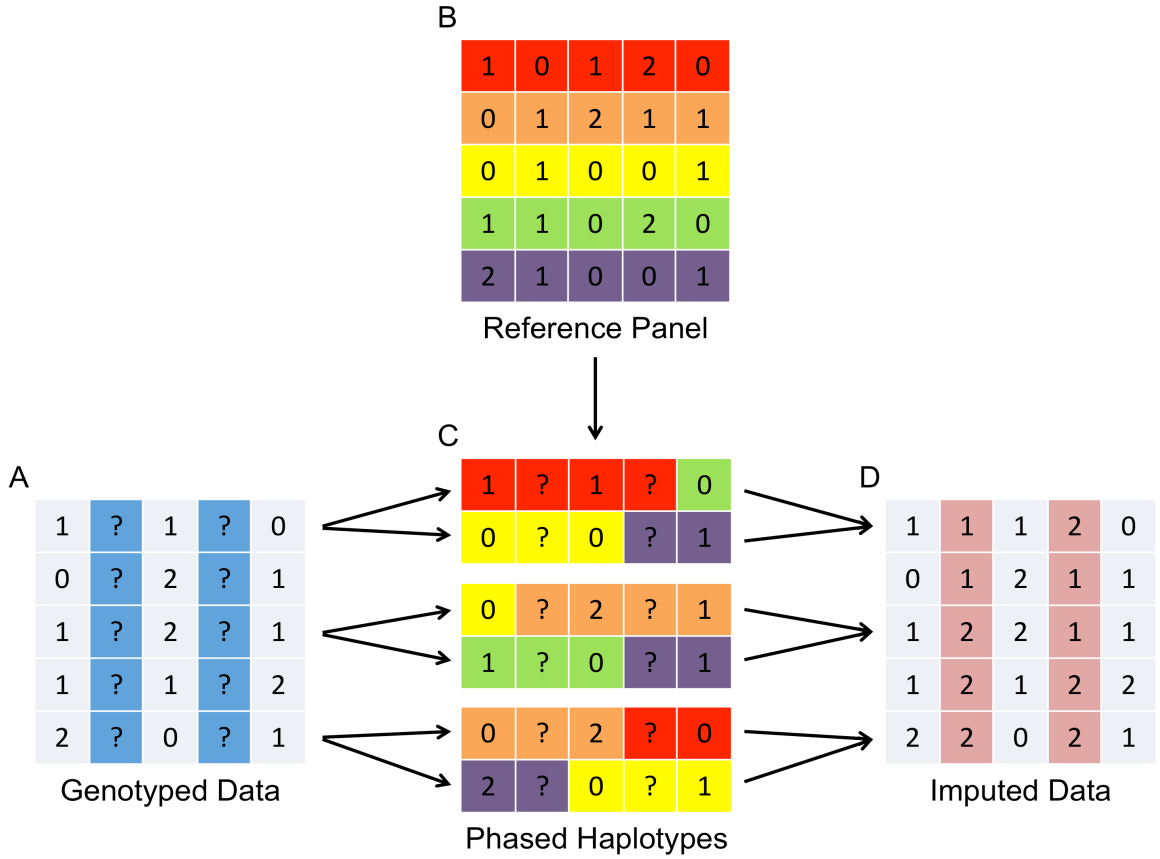

Figure 1.5: Genotype imputation. A: Genotyped data; missing data exists at untyped SNPs (shown by question marks, highlighted blue), B: Reference panel of haplotypes, C: Haplotypes from the genotyped data, phased with haplotypes from the reference panel, D: Imputed data, obtained from using the reference haplotypes to create the imputed genotypes (highlighted red) (adapted from Marchini and Howie, 2010).

Figure 1.5 shows how haplotype phasing (the estimation of missing haplotype information) through the use of a reference panel of fully sequenced SNPs can be used to obtain a set of imputed genotypes. However, Figure 1.5 is a somewhat simplified description of the process. In reality, there exists a level of uncertainty regarding the imputation of genotypes, and a probability distribution across all three possible genotypes is produced (Marchini and Howie, 2010). The commonly used imputation software IMPUTEv2 (Howie et al., 2009) reports this as an information metric known as info score, which takes a value between 0 and 1. Info score values close to 1 indicate a high certainty that a SNP has been imputed with the correct genotype (Zheng et al., 2015). Filtering for info score is often performed to remove poorly imputed SNPs from GWAS results. SNPs with an info score greater than 0.4 are often considered acceptable as well-imputed markers (Zheng et al., 2015), although high confidence in imputation fidelity has been described as an info score of above 0.8 (Huang et al., 2015). Genotype imputation can reveal many additional associations that are not possible to find using direct genotyping alone, and as such is now an essential tool for GWAS analyses (Li et al., 2009).

1.5.3 GWAS study design

1.5.3.1 Case-control and quantitative designs

There are two main types of phenotype that can be analysed using GWAS; categorical (usually a binary case/control phenotype), or quantitative. Statistically, quantitative traits are generally preferred as they have superior statistical power to detect genetic effects, and their outcomes are often easier to interpret (Bush and Moore, 2012).

1.5.3.2 Sample size and statistical power considerations

The selection of an appropriate sample size is a key element of GWAS study design. According to the CD/CV hypothesis, variants that contribute to complex traits are likely to have small effect sizes, which indicates that a large sample size is crucial if a GWAS is to be successful in identifying any significant associations (Hirschhorn and Daly, 2005). Another factor that increases the need for a large sample size is the number of tests that are performed during a GWAS. For a GWAS testing the association of 1,000,000 SNPs, 1,000,000 independent tests will be performed. Therefore, a correction for multiple testing must be introduced in order to minimise false positive associations. This leads to the implementation of a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$, which equates to $P < 0.05$ after using a Bonferroni correction for 1,000,000 independent tests (Risch and Merikangas, 1996). This genome-wide threshold is generally considered the de facto standard genome-wide threshold for GWAS analyses (Jannot et al., 2015). The Bonferroni correction is also the most conservative multiple testing adjustment, although other options, such as the false discovery rate (FDR) and permutation testing exist. However, the use of a less stringent P-value threshold would require follow-up studies to be performed in order to differentiate between false positives and genuine associations (Hirschhorn and Daly, 2005).

The choice of sample size is intrinsically linked to the statistical power of a study. Statistical power is defined as the probability of correctly rejecting the null hypothesis when a true association is present (or $1-\beta$, where β is the probability of a type II error) and is subject to factors outside of the control of an investigator, such as the effect sizes and MAFs of the underlying genetic variants, the history and genetic characteristics of the study population and the accuracy and completeness of the dataset. However, it is possible to maximise the statistical power of a study to ensure the best possible chance of obtaining meaningful results through careful selection of the study subjects, sample size, quality control (QC) methods and statistical analyses (Sham and Purcell, 2014).

These components of GWAS study design require careful consideration, as a combination of inadequate statistical power and an insufficiently stringent significance threshold can increase the number of false positive findings among significant results (Hoggart et al., 2008). If the power of a study is low, the genome-wide significance level has to be proportionally more stringent in order to maintain a fixed false positive report probability (FPRP).

Associations found to reach genome-wide significance from a more powerful study are more likely to represent true results than those from a low-powered study (Sham and Purcell, 2014). Generally, a GWAS with statistical power of above 80% is desired to be confident that any associations found are genuine. GWASs should always be thoughtfully designed, with careful power calculations determining the optimal choice of genotyping platform and sample size (Klein, 2007). Figure 1.6 shows the relationship between effect size, MAF and statistical power for a pre-specified sample size.

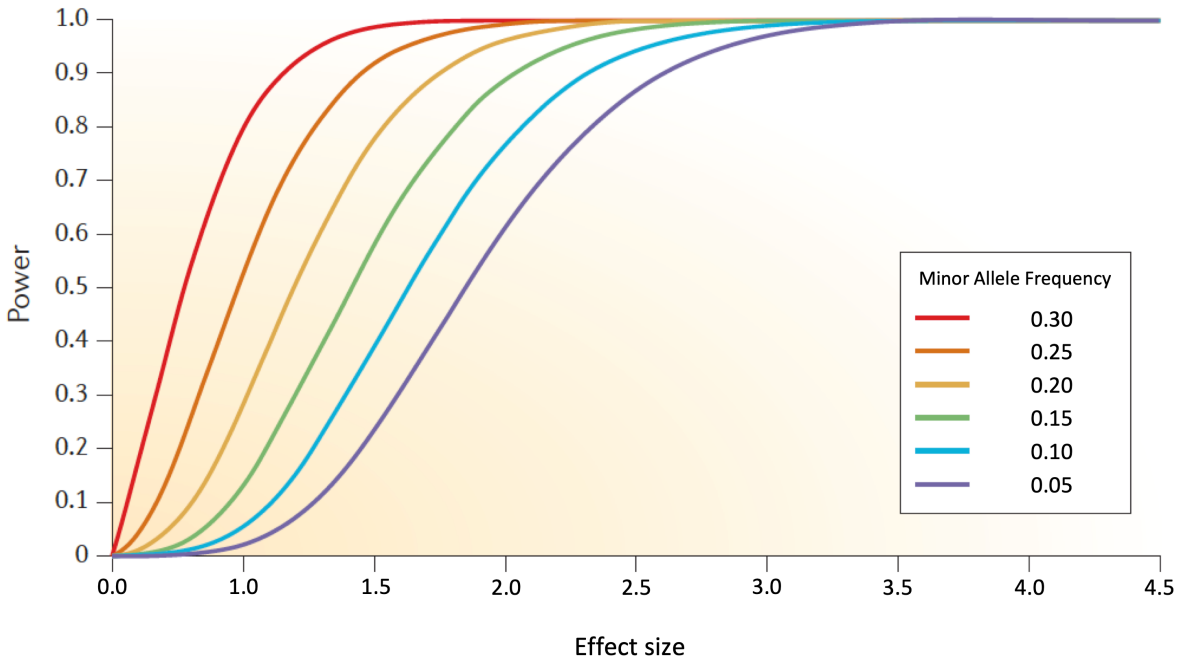

Figure 1.6: Example power curve. For a study with a pre-specified sample size, SNPs with a higher effect size will have greater statistical power. SNPs with higher minor allele frequency will also be associated with increased power (adapted from Sham and Purcell, 2014).

1.5.3.3 Underlying genetic analysis models

GWAS data is generally analysed through a series of single-locus statistical tests that examine each SNP independently for an association with a phenotype. The type of test conducted depends upon a variety of factors, and statistical tests vary for quantitative and case/control studies. Case/control studies usually employ contingency table methods or logistic regression, whereas quantitative traits are generally analysed using generalised linear model (GLM) approaches such as the analysis of variance (ANOVA), or Cox Proportional Hazards regression for survival data.

Regardless of the analysis method used and the trait type analysed, there are a variety of ways that genetic data can be encoded for association tests, which can have implications for the statistical power of the study (Bush and Moore, 2012). Allelic association tests analyse the association between one allele of a SNP and a phenotype, whereas genotypic association tests analyse the association between the phenotype and a genotype. The genotypes of different SNPs can also be classified into various genotypic models including additive, dominant and recessive models. Each of these models makes different assumptions regarding the genetic effect of the data, although the majority of GWASs tend to use an additive model. This is because the additive model has reasonable power to detect both additive and dominant effects, although it may be underpowered to detect some recessive effects. Multiple models can also be used to analyse genetic data, with the application of an appropriate correction for multiple testing (Bush and Moore, 2012).

1.5.3.4 Covariate adjustments

In addition to the selection of an appropriate genotypic encoding scheme, GWASs should be adjusted for factors that are known to influence the trait of interest (covariates), such as age, sex and any known clinical influences. Adjusting for these factors reduces the number of spurious associations due to sampling artifacts or inherent study design biases, although this may impact the statistical power of the study (Bush and Moore, 2012).

1.5.3.5 Population stratification

One of the most important covariates to consider when designing a GWAS is a measure of population stratification (otherwise known as population substructure) (Bush and Moore, 2012). Population stratification describes the existence of several subgroups within a population that differ in disease prevalence (or for quantitative traits, average trait value), which can cause systematic bias. If a genetic marker has different frequencies of occurrence within the subgroups, it can lead to false positive associations (Hirschhorn and Daly, 2005). The ethnicity of a subpopulation can have a dramatic difference in phenotype prevalence, and MAFs are variable across subpopulations. This can cause an issue when analysing cohorts consisting of multiple ethnicities, as SNPs specific to a particular ethnicity are likely to be associated to the phenotype of interest, thus skewing the results (Bush and Moore, 2012).

1.5.4 GWAS quality control

Inadequacies in study design and genotype calling errors can potentially introduce systematic biases into GWASs, which can cause an increase in both false positive and false negative associations (Anderson et al., 2010). QC of genetic data is an important part of a GWAS, which aims to minimise these potential false discoveries (Pongpanich et al., 2010). GWAS QC procedures are computationally intensive, operationally challenging and constantly evolving as best practices continue to be developed (Turner et al., 2011).

The following steps describe current standard procedures for GWAS QC, which are essential in order to be as confident as possible that the resulting associations are genuine.

1.5.4.1 SNP call rate filtering

The call rate of a SNP is the proportion of individuals in the study that do not have missing information for the corresponding SNP. Generally, the SNP call rate of a study is set at 95% (only SNPs with less than 5% missing information are analysed), although more stringent thresholds can be used, particularly in the case of smaller sample sizes (Reed et al., 2015).

1.5.4.2 Filtering for deviation from Hardy-Weinberg Equilibrium (HWE)

The Hardy-Weinberg principle explains how random mating produces and maintains a population with constant genotypic proportions, also known as Hardy-Weinberg Equilibrium (HWE) (Stark, 2015). Many GWASs tend to exclude SNPs that display significant deviation from HWE as this can be indicative of errors in genotyping or genotype calling (Anderson et al., 2010) or evidence of the presence of population substructure. While it is not always possible to ascertain which of these has occurred, it is common practice to assume a genotyping error and remove any SNPs violating HWE accordingly (Reed et al., 2015).

1.5.4.3 Minor Allele Frequency (MAF) filtering

As mentioned previously, inferring a statistically significant association between a SNP and a phenotype requires adequate statistical power. A considerable amount of homogeneity at a given SNP across patients in a sample cohort generally results in inadequate power with which to detect a significant association. This can occur when the MAF of a SNP is very small (i.e. the vast majority of patients in the cohort are homozygous for the major allele). It is for this reason that SNPs with low MAF are removed from GWAS analyses. The threshold at which MAF is filtered varies between studies, but generally a cut off of either 1% or 5% is applied (Reed et al., 2015).

Historically, genotyping platforms provide better coverage for SNPs with MAF of greater than 5% (Panagiotou et al., 2010), while higher statistical power has been reported for studies using a threshold of 5% (Fan et al., 2011). The majority of causal variants for common diseases identified through GWASs have also been shown to have a MAF greater than 5% in accordance with the CD/CV hypothesis (Bush and Moore, 2012). However, the use of a MAF threshold of 1% is more common for GWASs using genomic datasets with large enough sample sizes to ensure adequate statistical power, such as meta-analyses of several studies from multiple centres (Whiffin et al., 2014; Houlston et al., 2010). Regardless of the threshold chosen, the inclusion of MAF filtering is an integral step of GWAS QC.

1.5.4.4 Sample filtering

It is also important to remove any individuals from the study cohort that should be excluded from the analyses. Common criteria for sample filtering include outlying heterozygosity or missing genotype rates, duplicated or related individuals, discordant sex information and any individuals of divergent ancestry (Anderson et al., 2010). A frequently used measure of relatedness (or duplication) between pairs of samples is based on identity by descent (IBD), with an IBD kinship coefficient greater than 0.1 suggesting potential duplicates, relatedness or sample mixture (Reed et al., 2015). The individual of a related pair with the lowest genotype call rate is generally then removed. In cases where it is not possible to remove related individuals (i.e. siblings, cousins or parents), specially designed analysis techniques are required, such as mixed models regression analysis (Widmer et al., 2014).

1.5.5 Visualisation of GWAS results

The visualisation of GWAS results is a useful way of determining which, if any, SNPs are of statistical significance, as well as being a tool with which to identify any potential data inconsistencies or systematic biases that may have been overlooked during the study design and QC phases (Reed et al., 2015). There are two types of plot that are generally used to display GWAS results; Manhattan plots and quantile-quantile (Q-Q) plots.

1.5.5.1 Manhattan plots

The results of a GWAS are commonly visualised using a Manhattan plot (Figure 1.7A), which plots the -log10(P) of the associations for each SNP on the y-axis against their chromosomal position on the x-axis. Any regions that have numerous highly associated SNPs in LD appear as 'skyscrapers' (Turner, 2014). The genome-wide significance threshold ($P < 5.0\text{x}10^{-8}$) and the less stringent threshold of suggestive association ($P < 1.0\text{x}10^{-5}$) are also commonly displayed (the red and blue horizontal lines, respectively, Figure 1.7A). SNPs found above the threshold of suggestive association indicate a potential association that may require further investigation (Reed et al., 2015).

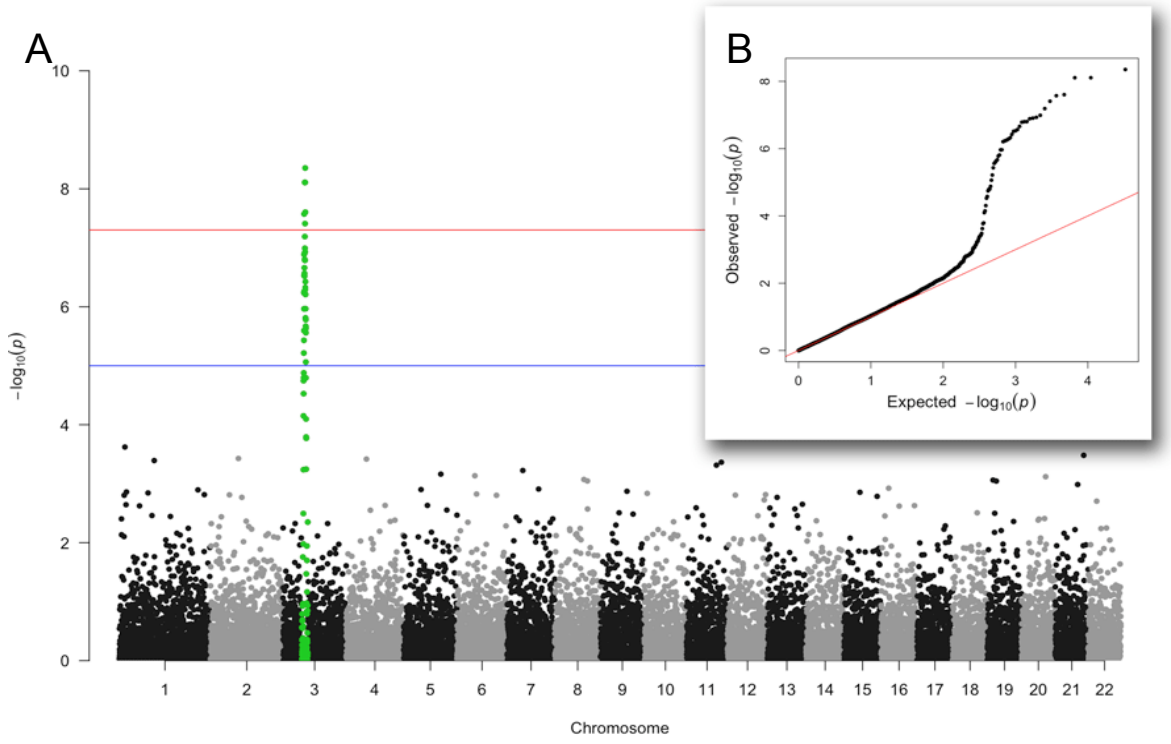

Figure 1.7: Example Manhattan and Q-Q plots. A: Manhattan plot showing -log10(P) of associated SNPs (y-axis) versus chromosomal location (x-axis). Red line: genome-wide significance threshold ($P < 5.0\text{x}10^{-8}$). Blue line: suggestive association threshold ($P < 1.0\text{x}10^{-5}$). SNPs of interest on chromosome 3 are highlighted in green. B: Q-Q plot showing the observed (y-axis) and expected (x-axis) -log10(P) values of associated SNPs. Substantial deviation from the diagonal (expected values) can be seen, indicating a potential issue with the dataset (adapted from Turner, 2014).

1.5.5.2 Quantile-Quantile (Q-Q) plots

Another useful visualisation tool for GWAS output is the Q-Q plot (Figure 1.7B). Q-Q plots show the observed P-values for all SNPs on the y-axis and the expected uniform distribution of P-values under the null hypothesis of no association on the x-axis (the diagonal line, Figure 1.7B). SNPs that exhibit strong associations will deviate from the diagonal at the upper right end of the plot. However, systematic deviation from the diagonal could indicate the presence of potential issues with the data such as cryptic relatedness or population stratification (Turner, 2014). Deviation from this line is measured using the λ statistic. A value of $\lambda \approx 1$ indicates that an appropriate adjustment for population substructure has been applied (Reed et al., 2015).

1.6 Further interrogation of GWAS results

1.6.1 Contextualising GWAS results using *in silico* resources

The identification of SNPs associated with a trait through GWAS analyses is far more useful when presented in a genetic context (Manolio, 2010). For example, it may be useful to know if a statistically significant SNP is intergenic, found within a protein-coding gene or near a methylation site in a specific cell type or tissue that is relevant to the disease being investigated (Reed et al., 2015). However, many of these SNPs are often not found within exonic genetic regions, but in intronic or intergenic regions (Manolio, 2010). This can make ascertaining which genes are affected, and how, extremely difficult. A wealth of *in silico* tools are available online that are designed to further investigate initial GWAS results. These are often free to use and can provide an insight into the genetic background and functional consequences of significantly associated SNPs identified through GWAS analyses.

1.6.1.1 Regional association analyses

When examining GWAS results, it is important to visually assess regions of the genome that harbour trait-associated loci in order to identify genes in the local region that may be impacted and ascertain the extent of LD between the variants identified (Pruim et al., 2010). LocusZoom (http://locuszoom.org) is a web-based plotting tool that enables the visualisation of GWAS results for a pre-specified region of the genome, which incorporates LD information from HapMap Phase II (Frazer et al., 2007) and the UCSC Genome Browser to identify and map SNPs of interest, and gives an overview of the extent of LD and the position of SNPs relative to nearby genes and areas of recombination (Pruim et al., 2010). An example of a regional association plot is shown in Figure 1.8.

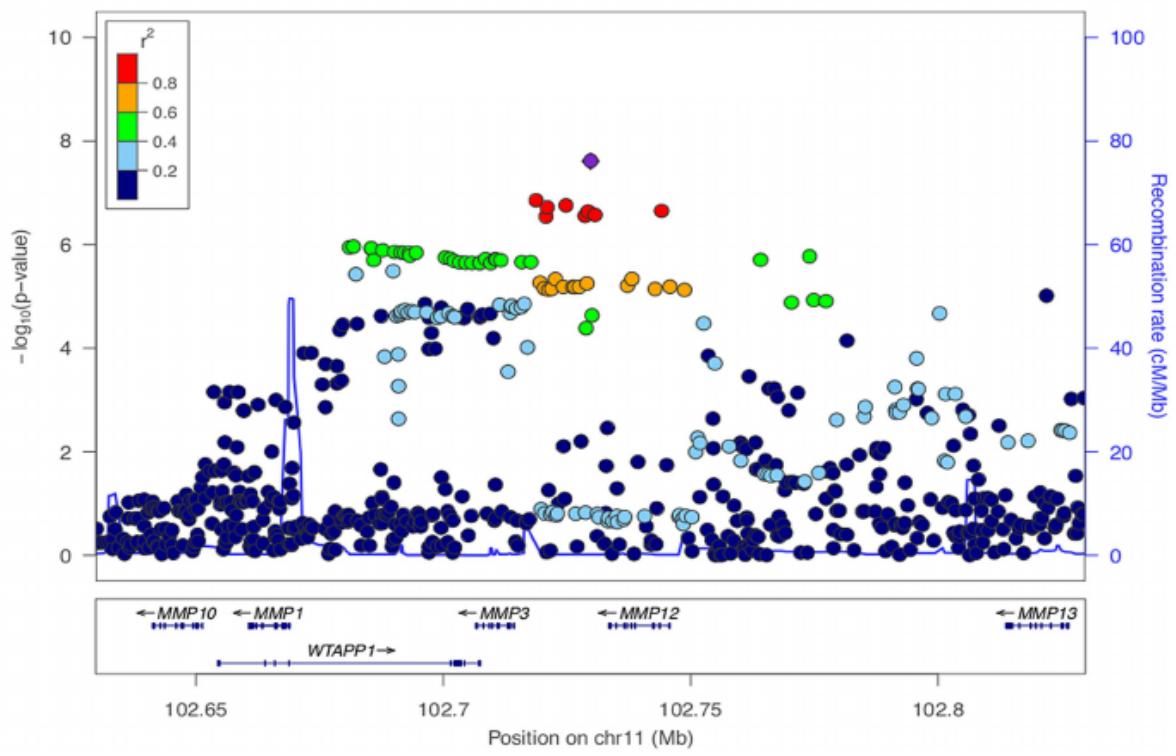

Figure 1.8: Example regional association plot. The -log10(*P*) of each SNP (y-axis) is plotted against the chromosomal position (x-axis). Each SNP is a colour-coded circle, the colour dependent upon the level of LD it shares with the reference SNP (the purple diamond; often chosen to be the most significant SNP). The key in the upper-left corner shows the LD thresholds for each SNP in terms of r^2. Genes in the region and their direction of transcription are shown in the rectangular box at the bottom of the plot (adapted from Traylor et al. 2014).

1.6.1.2 Expression quantitative trait loci (eQTL) analyses

Another potentially informative method of post-GWAS analysis is the investigation of eQTLs to identify causal genes of trait-associated SNPs and the functional mechanisms underlying these associations. An eQTL is a locus that accounts for part of the genetic variance of the expression phenotype of a gene (Nica and Dermitzakis, 2013), which can potentially aid the interpretation of GWAS findings through the identification of tissues that play a role in the pathogenesis of disease (Kang et al., 2012). There are two main types of eQTL; local (or cis-regulatory eQTLs) and distal (or trans-regulatory eQTLs) (Battle et al., 2017). A number of studies have utilised data from the Genotype-Tissue Expression (GTEx) Project (Carithers and Moore, 2015) to examine relationships between risk variants and gene expression (Loo et al., 2012; Hulur et al., 2015) and candidate susceptibility genes (Closa et al., 2014) for CRC. This database has also been utilised to conduct eQTL analyses for a number of cancers including CRC (Loo et al., 2017; Catalano et al., 2018) and breast cancer (Zhou et al., 2017a), as well as in other diseases including osteonecrosis of the jaw (Yang et al., 2018).

1.6.1.3 The PubMed database

The National Centre for Biotechnology Information (NCBI) online database PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) comprises of more than 30 million citations for biomedical literature from MEDLINE, life science journals and online books and can be utilised when contextualising the results of a GWAS to gain an understanding of the underlying biological mechanisms a SNP may impact, and can give an insight into the clinical relevance of the variant in question.

1.7 Validation of biomarkers through meta-analysis

In order for a biomarker to be considered for implementation in the clinical setting, it must firstly be validated in an independent cohort of patients, in accordance with the REporting recommendations for tumour MARKer prognostic studies (REMARK) guidelines (McShane et al., 2005). Validation cohorts often consist of a number of studies, which are combined to give a total sample size that achieves a pre-specified degree of statistical power to detect true associations between a variant and a trait of interest (Reed et al., 2015).

1.7.1 Visualisation of meta-analyses results

The results of meta-analyses are often visualised using forest plots. These plots show the individual effect sizes and confidence intervals for each study separately, and also provide an average effect size and confidence interval for all studies combined. Funnel plots are often used alongside forest plots in order to ascertain whether there is any underlying bias associated with the studies analysed. Smaller studies are likely to scatter widely at the bottom of the plot, with the spread narrower among larger studies (Sedgwick and Marston, 2015). Confidence intervals are often displayed as a dashed line, and in the absence of bias the plot will resemble a symmetrical, inverted funnel (Sedgwick and Marston, 2015). Examples of forest and funnel plots are shown in Figure 1.9.

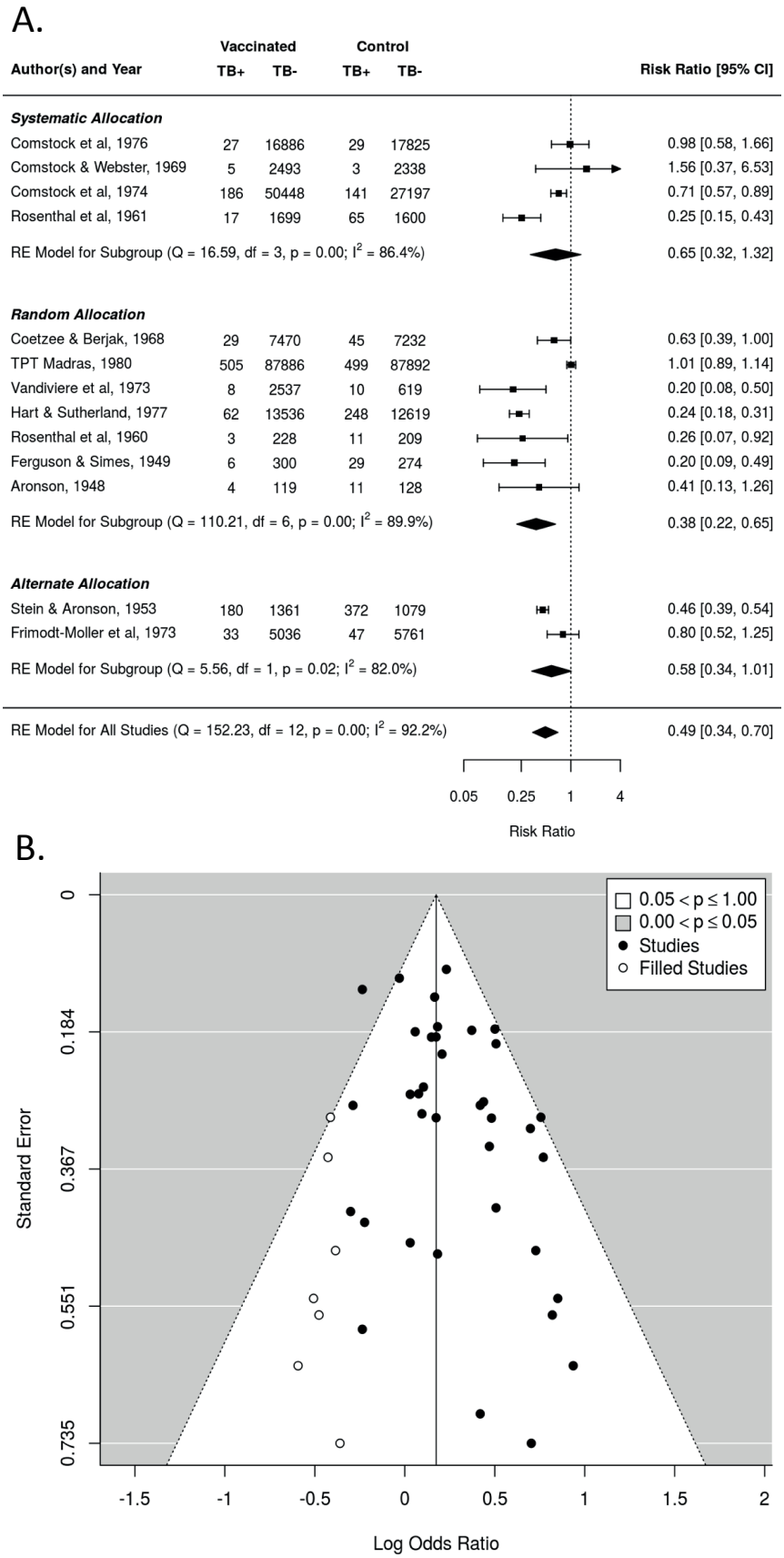

Figure 1.9: Example forest and funnel plots. A: Example of a forest plot. B: Example of a funnel plot (adapted from http://www.metafor-project.org/doku.php/metafor).

1.8 Hypothesis and aims of this project

1.8.1 Hypothesis

The hypothesis of this thesis was that novel common variants that influence the prognosis of patients with mCRC exist and are yet to be identified.

1.8.2 Aims of this thesis

- To study the influence of common somatic mutations on survival in mCRC
- To determine whether any novel germline variants are associated with survival in mCRC
- To identify possible underlying biological mechanisms that may be affected by these variants
- To attempt the validation of these variants as prognostic biomarkers for mCRC

Chapter 2

Materials and methods

The materials and methods described in this thesis are a combination of the candidate's own work and the work of others prior to the commencement of this project. For clarity, a flowchart highlighting all work undertaken both prior to and during this project is shown in Figure 2.1.

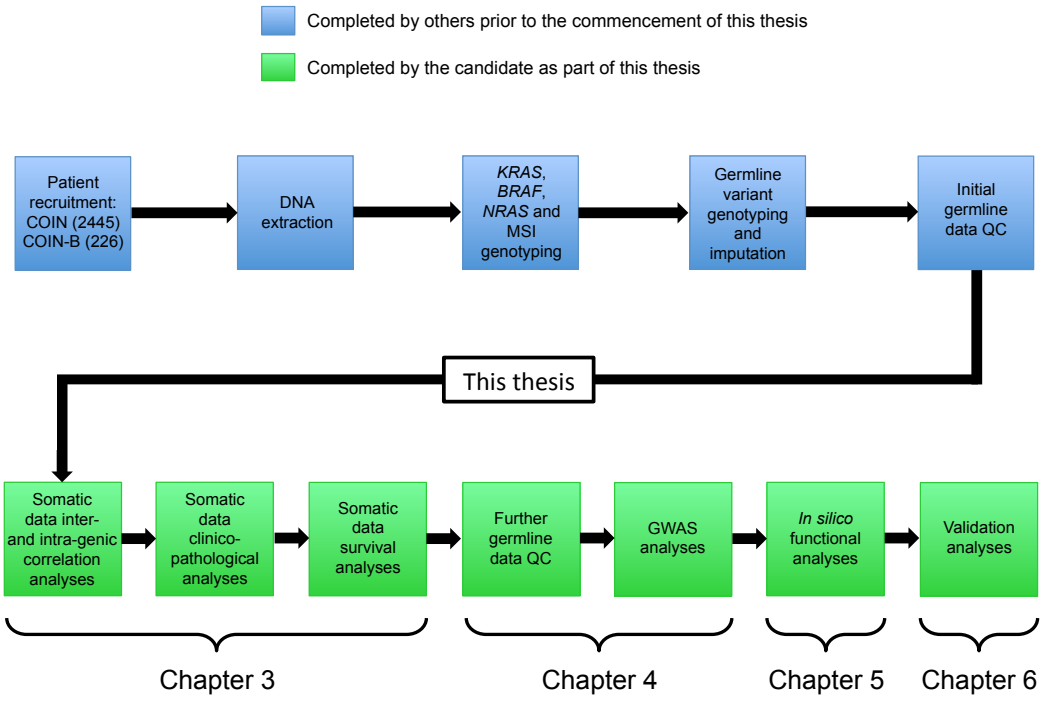

Figure 2.1: Analyses workflow. Workflow of analyses undertaken prior to (highlighted in blue) and during this project (highlighted in green). Patient recruitment to the COIN & COIN-B trials was conducted by the Medical Research Council's (MRC) Clinical Trials Unit (CTU). DNA extraction and analyses of patient tumour and germline samples was conducted as previously described (Smith et al., 2013; Al-Tassan et al., 2015). Initial QC of germline data was performed by Richard Houlston's laboratory at the Institute of Cancer Research (ICR) as previously described (Anderson et al., 2010). QC: Quality control. MSI: Microsatellite instability.

2.1 Patient characteristics

The patient data used in these analyses are from two Medical Research Council (MRC) clinical trials totalling 2671 patients with mCRC; COIN (2445 patients) and COIN-B (226 patients). Both trials received ethical approval by the respective ethical review boards (COIN: Medical Research and Ethics Committee (MREC): 04/MRE06/60, COIN-B: South West Multi-Centre Research Ethics Committee: 00316/0220/001-0001) and written informed consent was obtained from each participant, in accordance with the Declaration of Helsinki.

2.1.1 The COIN trial

The COIN trial was a three-armed randomised clinical trial in mCRC patients, funded by Cancer Research UK. Inclusion criteria comprised of written informed consent, age of at least 18 years, and with histologically confirmed primary adenocarcinoma of the colorectum, inoperable metastatic or locoregional measurable disease according to Response Evaluation Criteria In Solid Tumors (RECIST, version 1.0), no previous chemotherapy for metastatic disease, World Health Organization (WHO) performance status 0-2, and good end-organ function. Exclusion criteria comprised of previous or present malignant disease, uncontrolled medical comorbidity likely to interfere with COIN treatment or response assessment, known brain metastases, or previous oxaliplatin exposure (Adams et al., 2011).

COIN was designed to (I) assess the effect on OS of the addition of the EGFR-targeted antibody cetuximab to oxaliplatin-based first-line continuous chemotherapy for mCRC, and (II) determine whether intermittent palliative chemotherapy resulted in non-inferiority in terms of OS when compared to continuous chemotherapy, irrespective of *KRAS* mutation status. Patients were recruited between 9th March 2005 and 9th May 2008 by consultant oncologists at 111 centres across the UK and Republic of Ireland that routinely undertake treatment of mCRC, and randomly assigned to one of three arms at a ratio of 1:1:1. Arm A received continuous oxaliplatin-based chemotherapy, Arm B received continuous chemotherapy plus cetuximab and Arm C received intermittent chemotherapy (Figure 2.2). The choice of chemotherapy to partner oxaliplatin was made by all patients prior to randomisation. Two thirds of patients chose oral capecitabine, with one third choosing infusional fluoropyrimidine (Smith et al., 2013). Patients in Arms A and B continued treatment until either the development of progressive disease or cumulative toxic effects, or until the patient decided to stop (Adams et al., 2011).

The results of COIN did not confirm a benefit of addition of cetuximab to oxaliplatin-based chemotherapy in first-line treatment of mCRC patients. There also was no evidence of benefit in PFS or OS in *KRAS* wild type patients, or in patients selected by additional mutational analysis of their tumours (Maughan et al., 2011). Therefore, the use of cetuximab in combination with oxaliplatin and capecitabine in first-line chemotherapy in mCRC patients could not be recommended. COIN did not show non-inferiority of intermittent compared with continuous chemotherapy for mCRC in terms of OS (Adams et al., 2011). The trial is registered; ISRCTN27286448.

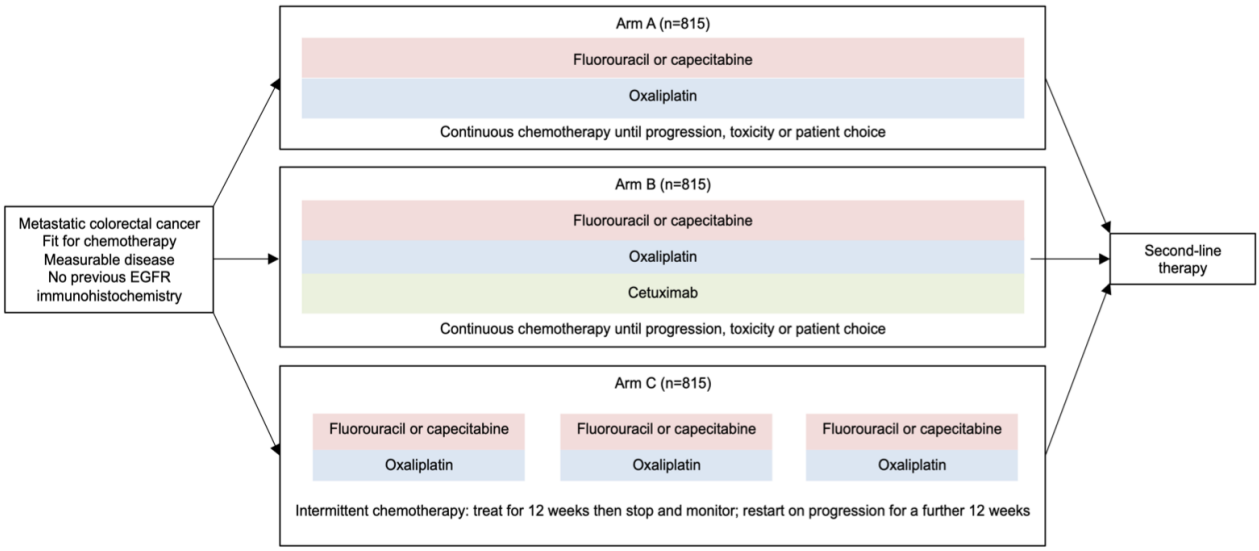

Figure 2.2: COIN trial design. Adapted from Adams et al., 2011.

2.1.2 The COIN-B trial

The COIN-B trial was an adjunct to COIN, designed to establish how cetuximab might be safely and effectively added to intermittent chemotherapy. COIN-B was a two-armed randomised clinical trial in mCRC patients, funded by the MRC and Merck KGaA. Inclusion and exclusion criteria for COIN-B was identical to that of COIN, except that inoperable metastatic or locoregional measurable disease was defined according to RECIST version 1.1 (Wasan et al., 2014). Patients were recruited between 13th July 2007 and 6th March 2010 from 30 centres across the UK and one in Cyprus, and randomised at a ratio of 1:1. Arm D received intermittent chemotherapy plus intermittent cetuximab and Arm E received intermittent chemotherapy with continuous cetuximab. Both groups received FOLFOX and weekly cetuximab for 12 weeks, followed by planned interruption (Arm D) or planned maintenance (Arm E) (Figure 2.3).

COIN-B was suspended in May 2008 as emerging data showed that *KRAS* mutations were predictors to EGFR-targeted moAB treatment (Lievre et al., 2006; Amado et al., 2008; Karapetis et al., 2008). The trial restarted in January 2009 and included prospective *KRAS* mutation analysis prior to randomisation. Only patients with *KRAS* wild type tumours were recruited following the reactivation of the trial. The *KRAS* mutation status of previously enrolled patients was assessed retrospectively while the trial was suspended. The results of COIN-B showed that cetuximab could be safely incorporated into two first-line intermittent chemotherapy strategies (Wasan et al., 2014). The trial is registered; ISRCTN38375681.

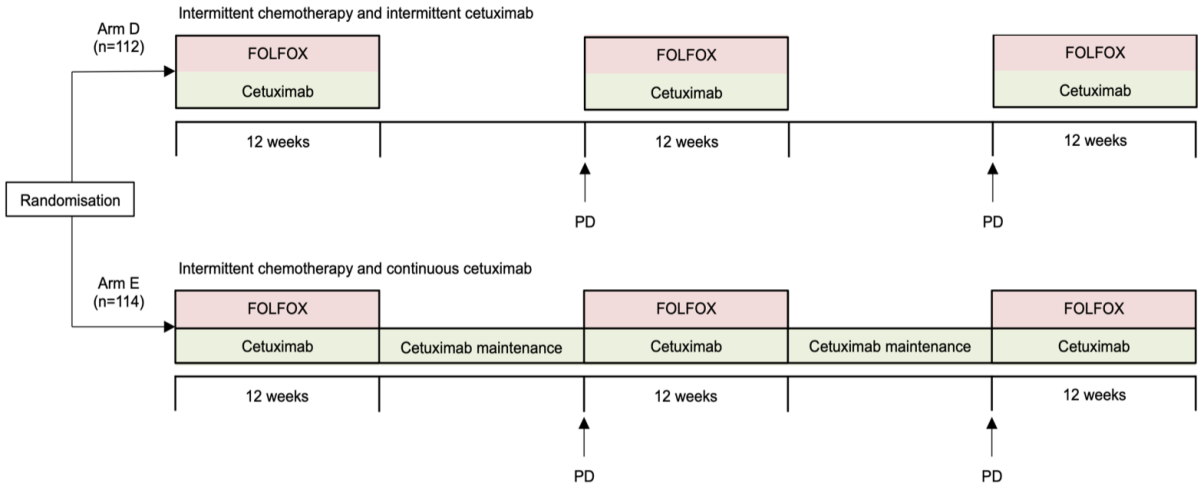

Figure 2.3: COIN-B trial design. Adapted from Wasan et al., 2014. Treatment cycles continued until progressive disease (PD) with maximally tolerated treatment, or patient choice. FOLFOX: Folinic acid and oxaliplatin followed by bolus and infused fluorouracil (also known as Oxaliplatin modified de Gramont; OxMdG).

Clinicopathological data for patients in COIN and COIN-B are described in Table 2.1.

Table 2.1: Clinicopathological data for patients in COIN & COIN-B.

Trial	**COIN**			**COIN-B**	
Arm	**A**	**B**	**C**	**D**	**E**
Number of patients	815 (30.5)	815 (30.5)	815 (30.5)	112 (4.2)	114 (4.3)
Number genotyped	651 (80)	674 (83)	661 (81)	106 (95)	111 (97)
Age					
Mean (SD)	61.9 (10)	62.8 (10)	62.5 (10)	62.5 (10)	62.2 (11)
$<$20	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)
20-49	87 (11)	88 (11)	77 (10)	14 (13)	14 (12)
50-59	212 (26)	172 (21)	194 (24)	17 (15)	28 (25)
60-69	327 (40)	341 (42)	330 (41)	52 (46)	40 (35)
70-79	182 (22)	206 (25)	210 (26)	26 (23)	30 (26)
80-89	7 (1)	8 (1)	3 (0)	3 (3)	2 (2)
Male	525 (64)	543 (67)	523 (64)	66 (59)	65 (57)
Female	290 (36)	272 (33)	292 (36)	46 (41)	49 (43)
KRAS mutant	267 (41)	294 (44)	258 (39)	24 (23)	15 (14)
KRAS wild type	367 (56)	362 (54)	396 (60)	78 (74)	91 (82)
BRAF mutant	57 (9)	45 (7)	73 (11)	8 (8)	16 (14)
BRAF wild type	575 (88)	617 (92)	579 (88)	60 (57)	67 (59)
NRAS mutant	18 (3)	31 (5)	19 (3)	7 (7)	8 (7)
NRAS wild type	613 (94)	627 (93)	630 (95)	62 (59)	76 (69)
MSI	19 (3)	26 (4)	21 (3)	NA	NA
MSS	483 (74)	494 (73)	522 (79)	2 (2)	NA
OxMdG	279 (34)	281 (35)	282 (35)	112 (100)	114 (100)
XELOX	536 (66)	534 (66)	533 (65)	0 (0)	0 (0)
Cetuximab	0 (0)	815 (100)	0 (0)	112 (100)	114 (100)
No Cetuximab	815 (100)	0 (0)	815 (100)	0 (0)	0 (0)
Mean OS (days)	487	478	468	508	457
Median OS (days)	480	464	436	478	430
Complete response	40 (5)	50 (6)	22 (3)	8 (7)	7 (6)
Missing result	101 (12)	98 (12)	82 (10)	10 (9)	18 (16)
Partial response	377 (46)	383 (47)	399 (49)	69 (62)	53 (47)
Progressive disease	113 (14)	113 (14)	112 (14)	10 (9)	15 (13)
Stable disease	184 (23)	171 (21)	200 (25)	15 (13)	21 (18)

Uncharacterised mutants and failed genotyping data omitted. Percentages are shown in parentheses unless otherwise stated. Age: Age at randomisation. SD: Standard deviation. OS: Overall survival. NA: Not assessed.

2.2 Specimen characteristics and assay methods

As shown in Figure 2.1, all methods described in this section were performed by others as previously described (Anderson et al., 2010; Smith et al., 2013; Al-Tassan et al., 2015).

2.2.1 Somatic tumour DNA analyses

Somatic *KRAS*, *BRAF*, *NRAS* and MSI status of patient tumour samples was determined as previously described (Maughan et al., 2011; Smith et al., 2013). Somatic mutations were genotyped using a combination of two mutation detection platforms; pyrosequencing and Sequenom. Somatic mutations were screened for in *KRAS* (codons 12, 13 and 61) and *BRAF* (codon 600) using both pyrosequencing and Sequenom, and in *BRAF* (codon 594) and *NRAS* (codons 12 and 61) using Sequenom only. Both technologies had high genotype success rates; 41,944/43,340 (96.8%) for Sequenom, and 21,016/25,200 (83.4%) for pyrosequencing. For samples analysed by both technologies (n=1,612), genotype concordance in *KRAS* was 99.1% (8,642/8,719 calls were concordant). Sanger sequencing was used to infer the genotypes of discordant samples between pyrosequencing and Sequenom. For remaining calls where Sanger sequencing failed, the mutant genotype was selected due to there being an obvious mutant trace via one technology. MSI status was determined using the markers BAT-25 and BAT-26 (Smith et al., 2013). Tumour DNA samples were extracted from FFPE and were available from 2184 cases (1976 from COIN and 208 from COIN-B) (Smith et al., 2013; Wasan et al., 2014).

2.2.2 Germline DNA analyses

All genotyping of germline samples was performed at the Department of Genetics, King Faisal Specialist Hospital and Research Centre (Riyadh, Saudi Arabia), as previously described (Al-Tassan et al., 2015). Cases were genotyped using Affymetrix Axiom Arrays according to the manufacturer's recommendations (Santa Clara, USA), using duplicate samples and sequencing of significantly associated SNPs in a subset of samples to confirm genotyping accuracy. For all variants $>$ 99% concordant results were obtained (Al-Tassan et al., 2015).

All initial QC and imputation of germline data was performed by Professor Richard Houlston's laboratory at the Institute of Cancer Research (ICR) as previously described (Anderson et al., 2010; Al-Tassan et al., 2015). Initially, blood DNA samples were available from 2244 patients. Of these cases, a proportion of individuals were excluded that had $<$ 95% successfully genotyped SNPs (n=122), discordant sex information (n=8), classified as out of bounds by Affymetrix (n=30), duplication or crypted relatedness (IBD $>$ 0.185, n=4) and evidence of non-white European ancestry (n=130). Variants with call rate $<$ 95%, departure from HWE at $P < 1.0\text{x}10^{-5}$ and MAF $<$ 0.01 were excluded.

Phasing of variant genotypes was performed using SHAPEIT (https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html) and prediction of the untyped SNPs was performed using IMPUTE v2.3.0 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html) based on the data

from the 1000 Genomes Project (Phase 1 integrated variant set, v3.20101123) (https://www.internationalgenome.org) as the reference panel. Concordance between sequenced and imputed variants was assessed to estimate imputational fidelity in a subset of cases (n=200). Following genotyping and imputation, data for 1950 CRC cases were available for analyses (Al-Tassan et al., 2015). Two patients had missing survival data and were subsequently removed, thus 1948 patients were available for analysis (1778 from COIN and 170 from COIN-B).

2.3 Data files

The somatic and germline data resulting from the analyses in Chapter 2.2 were obtained by the candidate as (I) a Microsoft Excel file (for somatic tumour sample data and other clinical factors of the COIN and COIN-B cohorts) and (II) 22 Oxford text genotype (.gen) files (for germline data) prior to the commencement of this project. It is these files that were analysed by the candidate, the results of which are described in this thesis (Figure 2.1).

2.3.1 Clinical molecular data

Clinical molecular data for all patients in COIN and COIN-B was obtained in Microsoft Excel spreadsheet format, and contained the data shown in Table 2.2.

Table 2.2: Description of COIN & COIN-B clinical molecular patient data.

Name	Description	Name	Description
patid	Patient ID	SEX	Sex of patient
AGE	Age (at randomisation)	age65	Age over 65yrs (yes/no)
trial	Trial (COIN/COIN-B)	TRT	Treatment arm
CHEMO	Chemotherapy regimen	PRT	Radiotherapy (yes/no)
CREATC	Creatinine clearance level	SA	Surface area of tumour
DODIAG	Date of diagnosis	DODM	Date of metastatic disease diagnosis
DOR	Date of randomisation	dls	Date last seen
DOD	Date of death	death	Patient died (yes/no)
KRAS	*KRAS* mutation status	BRAF	*BRAF* mutation status
NRAS	*NRAS* mutation status	MSI	MSI status
prog	Progressed yes/no	ADJCH	Prior adjuvant chemotherapy (yes/no)
progdate	Date of first progression	MLIV	Liver metastases (yes/no)
MLNG	Lung metastases (yes/no)	MNODE	Nodal metastases (yes/no)
MPERI	Peritoneal metastases (yes/no)	MOTH	Other metastases (yes/no)
metscat	Synchronous vs metachronous metastases	metsites	Number of metastatic sites
diagtime	Time since diagnosis (at randomisation)	metstime	Time from primary diagnosis to metastases
resp12	Response at 12 weeks (yes/no)	dstat12	Response status at 12 weeks
bestresp	Any response on trial (yes/no)	beststat	Best response status on trial
TSTAT	Resection status of primary tumour	colon	Primary tumour in colon (yes/no)
SITEPT	Site of primary tumour	SPSITE	Free-text description of primary tumour
SVOM	Worst 12 week CTC: Vomiting	SNAUS	Worst 12 week CTC: Nausea
SSTOM	Worst 12 week CTC: Stomatitis	SNAILC	Worst 12 week CTC: Nail changes
SDIAR	Worst 12 week CTC: Diarrhoea	SLETH	Worst 12 week CTC: Lethargy
SNEUT	Worst 12 week CTC: Neutrophils	SRASH	Worst 12 week CTC: Skin rash
SHFS	Worst 12 week CTC: Hand-foot syndrome	SPNP	Worst 12 week CTC: Peripheral neuropathy

SHYPMG	Worst 12 week CTC: Hypomagnesaemia	SPNP24	Worst 24 week CTC: Peripheral neuropathy
neutsep	Worst 12 week CTC: Neutropenic sepsis	WHO	WHO PS
ALKP	Alkaline phosphatase level	alkp	ALKP (grouped)
CEA	Carcinoembryonic antigen level	cea	CEA (grouped)
GFR	Glomerular Filtration Rate	gfr	GFR (grouped)
PLT	Platelet count	plt	PLT (grouped)
WBC	White blood cell count	wbc	WBC (grouped)

CTC: Cytotoxicity. MSI: Microsatellite instability. PS: Performance status.

2.3.2 Genomic data

Genomic data was obtained in Oxford text genotype (.gen) file format. Twenty-two files were obtained; one for each autosome. This file type contains one row per variant. The first five columns contain information on chromosome, variant ID, base pair coordinate, allele 1 (usually minor) and allele 2 (usually major). Each subsequent triplet of values indicates the likelihoods of homozygote A1, heterozygote and homozygote A2 genotypes at this variant, respectively, for each member of the sample.

2.4 Analysis hardware

All analyses were performed using an Apple (Cupertino, USA) MacBook Pro (Retina, 15-inch, Mid 2014, 2.8GHz Intel Core i7 processor, 16GB 1600 MHz DDR3), running operating system OS X Yosemite. For analyses requiring intensive computation (namely conversions of large data files and multivariable GWAS analyses), Cardiff University's high-performance cluster (HPC) Raven was used via command line-based remote access. Advanced Research Computing at Cardiff (ARCCA) granted access to the HPC for these analyses.

2.5 Analysis software

2.5.1 R 3.5.2

R is an open source language and environment for statistical computing and graphics, freely available to download from http://www.r-project.org, and was used for the vast majority of analyses performed as part of this project.

2.5.1.1 RStudio 1.0.153

R was used in conjunction with the integrated development environment (IDE) RStudio. RStudio is freely available to download from https://www.rstudio.com/products/RStudio/Desktop.

2.5.1.2 R Packages

R packages are available for download and installation within the R environment from the Comprehensive R Archive Network (CRAN) repository. All R packages used in this project are listed in Table 2.3.

2.5.1.2.1 R package: survival

survival is an R package containing all the core survival analysis routines including Kaplan-Meier curves and Cox models, and was used extensively for all survival analyses.

2.5.1.2.2 R package: GenABEL

GenABEL is an R library for GWAS analysis (Aulchenko et al., 2007), and was used extensively for all GWAS analyses. GenABEL provides an efficient file format for storing genotype data and enables pre-GWAS QC, while also being a tool for running GWASs of both continuous (quantitative) and binary (case/control) phenotypes (Karssen et al., 2016).

Table 2.3: R software packages utilised in this project.

R Package	Description	Function	Description
base	The R base package: a set of functions automatically loaded in R	cbind	Combines R objects by columns
		colnames	Retrieves or sets the column names of a matrix-like object
		length	Returns or sets the length of vectors, Lists and factors
		library	Loads package into workspace
		nrow	Returns the number of rows of an object
		order	Returns a permutation which rearranges its first argument into ascending or descending order
		rbind	Combines R objects by rows
		subset	Returns subsets of vectors, matrices or data frames that meet specified conditions
		summary	Produces object summaries
car	Companion to applied regression	recode	Recodes a variable
corrplot	Visualisation of a correlation matrix	corrplot	Visualisation of a correlation matrix
GenABEL	Genome-wide SNP association analysis	convert.snp.tped	Converts genotypic data in transposed-ped format (.tped and .tfam) to internal genotypic data formatted file
		GASurv	Makes survival object for use with mlreg
		load.gwaa.data	Loads genotype and phenotype data from files to gwaa.data object

		mlreg	Linear and logistic regression and Cox models for genome-wide SNP data
ggplot2	Creates data visualisations using the grammar of graphics	aes	Constructs aesthetic mappings
		ggplot	Creates a new ggplot
		theme	Modifies components of a theme
Hmisc	Contains functions including some useful for data analysis, high-level graphics and utility operations	rcorr	Computes a matrix of correlation coefficients for all possible pairs of columns of a matrix
meta	General package for meta-analysis	forest	Forest plot displaying results of meta-analysis
		funnel	Funnel plot for assessing small study effects in meta-analysis
		metagen	Generic inverse variance meta-analysis
		summary	Displays summary statistics for meta-analysis including the Cochrane's Q and I^2 Tests of heterogeneity
qqman	Q-Q and Manhattan plots for GWAS data	manhattan	Creates a Manhattan plot from GWAS results
		qq	Creates a Q-Q plot from p-values from a GWAS study
RColorBrewer	ColorBrewer palettes for use with graphic visualisations of data	brewer.pal	A selection of ColorBrewer palettes
stats	The R base statistics package: a set of functions automatically loaded in R for statistics	chisq.test	Calculates Chi-squared test
		cor	Calculates the correlation of x and y
		fisher.test	Calculates Fisher's exact test
survival	Survival analysis	coxph	Fits Cox proportional hazards regression model

		print.survfit	Returns a short summary of a survival curve
		summary.survfit	Returns a list containing the survival curve with confidence limits and other information
		Surv	Creates a survival object
		survdiff	Tests whether there is a significant difference between two or more survival curves.
		survfit	Creates survival curves from a formula, a previously fitted Cox model, or a previously fitted accelerated failure time model
survminer	Creates survival curves using ggplot2	ggsurvplot	Creates survival curves using ggplot2
utils	R utility functions	head	Returns the first six lines of a vector, matrix, table, data frame or function
		install.packages	Downloads and installs packages from CRAN repositories or from local files
		tail	Returns the last six lines of a vector, matrix, table, data frame or function

2.5.2 Cyberduck

Cyberduck is a free file transfer assistant and cloud storage browser available to download from https://cyberduck.io. Here, it was used to transfer files between folders on the local machine and folders within the Raven server for file conversions requiring a large amount of computational power.

2.5.3 GTOOL

GTOOL is a command line program for transforming sets of genotype data, available to download from http://www.well.ox.ac.uk/cfreeman/software/gwas/gtool.htmlDownload. GTOOL was used in conjunction with Raven to convert germline data files from .gen to .ped format.

2.5.4 PLINK 1.9

PLINK is a command line based toolset specifically designed for whole genome association analysis, which can be downloaded from https://www.cog-genomics.org/plink2. PLINK 1.9 was used for multiple file conversions, MAF filtering and all LD analyses.

2.5.5 SNPTEST

SNPTEST is a program for the analysis of single SNP association in genome-wide studies, freely available to download from https://mathgen.stats.ox.ac.uk/genetics_ software/snptest/ snptest.htmldownload. SNPTEST was used to obtain info score data for all analysed variants.

2.5.6 LocusZoom

LocusZoom is a suite of tools to provide fast visualisation of SNPs found through GWAS analyses, which includes LD information. It can be accessed in online and stand-alone form. Here, the online version of the software was used for all regional association plots, available at http://locuszoom.org/genform.php?type=yourdata.

2.5.7 The GTEx Project database

The GTEx Project database, accessible at https://gtexportal.org/home/, is a comprehensive database of gene expression data for a range of tissues. Genetic associations between variants of interest and their potential impact on gene expression for a range of tissues can be identified and multi-tissue eQTL plots produced using the in-built plotting function.

2.5.8 The PubMed database

The NCBI online database PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) comprises of more than 30 million citations for biomedical literature from MEDLINE, life science journals and online books.

2.5.9 Microsoft Excel 2011

Microsoft Excel was used for the tabulation of power calculation data prior to this information being plotted in power curves.

2.5.10 Microsoft PowerPoint 2011

Microsoft PowerPoint was used to edit some figures prior to their inclusion in this thesis.

2.5.11 LaTeX

LaTeX is a document preparation system, which was used to typeset this thesis. It is freely available to download from https://www.latex-project.org.

2.6 Study design and statistical analysis methods

All analyses in this thesis are retrospective. No stratification by disease stage was employed (due to all patients included having Stage IV disease). The primary endpoint was OS; the time from trial randomisation to death. The Bonferroni correction for multiple testing was used throughout this thesis as it is the most conservative and least computationally intensive method, although alternative methods are available including FDR and permutation testing (Bush and Moore, 2012). All P-values presented in this thesis are uncorrected unless otherwise stated. Sample size was predetermined by the number of patients in the COIN and COIN-B trials that passed QC and had available survival data (patients with missing survival data were not included in these analyses). Study design and statistical analysis methods relating to individual chapters can be found in Chapters 3.2.2, 4.2.2, 5.2.1 and 6.2.2, respectively.

Chapter 3

Inter-relationships between somatic mutations and their influence on survival in mCRC

Results from this chapter were published in:
Matthew G. Summers, Christopher G. Smith, Timothy S. Maughan, Richard Kaplan, Valentina Escott-Price and Jeremy P. Cheadle. *BRAF* and *NRAS* Locus-Specific Variants Have Different Outcomes on Survival to Colorectal Cancer. *Clinical Cancer Research* (2017), 23(11); 2742-9. DOI: 10.1158/1078-0432.CCR-16-1541.

3.1 Introduction

Despite advances in CRC screening and treatment in recent years and the subsequent increase in survival rates in many countries, CRC remains the fourth largest cause of cancer-related death worldwide (Brenner et al., 2014). Furthermore, the five-year survival of mCRC patients is drastically lower than that of patients with locally advanced disease (Miller et al., 2019). A number of factors that influence CRC prognosis have been suggested, however, clinically relevant biomarkers are limited (McShane et al., 2005). A clear need exists to identify prognostic biomarkers that may help clinicians in the management of patients with CRC.

The main prognostic marker used in clinical practice after diagnosis of CRC is clinicopathological staging (Walther et al., 2009), although additional markers are recommended for prognostic assessment, including MSI status in Stage II (Labianca et al., 2013) and *BRAF* mutation status in Stage IV disease (Van Cutsem et al., 2016). Other factors identified as having an influence on CRC prognosis are systemic inflammatory response to the tumour (Leitch et al., 2007) and the type, density and location of immune cells in the tumour (Galon et al., 2006), while lifestyle factors such as physical activity, body mass index (BMI) and smoking have also been shown to influence CRC prognosis (Haydon et al., 2006; Reeves et al., 2007; Boyle et al., 2013). In the case of mCRC, Köhne's prognostic classification, based on PS, WBC count, ALKP levels and number of metastatic sites has been proposed (Kohne et al., 2002), with PS being a strong prognostic and predictive factor for chemotherapy (Van Cutsem et al., 2016).

The molecular pathogenesis of CRC is heterogeneous, and the mechanisms underlying CRC development are clinically important because they are linked to response to treatment (Sadanandam et al., 2013) and patient prognosis (De Sousa et al., 2013). The most commonly studied genetic makers of CRC prognosis have historically been somatic markers associated with tumour progression in the adenoma-carcinoma sequence or genomic instability, for which some biological rationale exists for a potential effect on prognosis (Walther et al., 2009).

Some germline markers have also been shown to influence prognosis (Smith et al., 2015; Phipps et al., 2016); these are investigated in Chapter 4.

Since the role of mutations in *KRAS* as predictors of the effect of EGFR-receptor blocking therapy in CRC was identified (Lievre et al., 2006; Amado et al., 2008; Karapetis et al., 2008), mutations within the EGFR signalling pathway have been studied further in order to evaluate their prognostic role in CRC. It has been shown that *KRAS*, *BRAF* and *NRAS* mutations are associated with poor prognosis (Richman et al., 2009; Eklof et al., 2013; Schirripa et al., 2015). Presence of MSI has also been shown to confer an inferior prognosis in mCRC patients (Tran et al., 2011; Smith et al., 2013), in contrast to the positive impact on survival reported for patients with locally advanced disease (Popat et al., 2005; Benatti et al., 2005; Bertagnolli et al., 2009; Hutchins et al., 2011). The inferior prognosis of mCRC patients with MSI-positive tumours has been suggested to be driven by the association of MSI with *BRAF* mutations (Tran et al., 2011), although MSI has also been identified as an indicator of poor prognosis when analysed independently of *BRAF* mutation status (Smith et al., 2013). Similarly, the superior prognosis of MSI in Stage II patients has also been identified in patients with both *BRAF* mutant and *BRAF* wild type tumours (Lochhead et al., 2013). Furthermore, there is growing evidence that CIMP-driven epigenetic alterations characterise a subgroup of CRCs with a distinct aetiology and prognosis (Juo et al., 2014), although this has not yet been recommended for testing in clinical practice (Van Cutsem et al., 2016).

In this chapter, the frequency of somatic mutations within three genes involved in the EGFR signalling cascade that are known to influence CRC prognosis (*KRAS*, *BRAF* and *NRAS*) and the MSI status of tumour specimens collected from patients with mCRC in the COIN and COIN-B trials were examined, their inter- and intra-genic mutation correlations analysed and clinicopathological characteristics investigated. This was followed by survival analyses of the COIN and COIN-B patient cohorts in order to investigate in more detail the prognostic value of these markers, and whether any further associations with survival could be identified.

3.1.1 Aims and objectives

The aims and objectives of this chapter were as follows:

- To identify the inter- and intra-genic relationships between somatic mutations and MSI status to account for underlying prognostic effects in survival analyses
- To analyse clinicopathological characteristics according to somatic mutation and MSI status
- To assess whether combining the COIN and COIN-B datasets for survival analyses was feasible
- To perform univariable and multivariable survival analyses stratified by somatic mutation and MSI status

3.2 Materials and methods

3.2.1 Patient and specimen characteristics

Somatic tumour DNA samples were available from 2184 patients in the MRC clinical trials COIN (Chapter 2.1.1; n=1976) and COIN-B (Chapter 2.1.2; n=208). COIN and COIN-B tumour samples were collected as FFPE blocks (Smith et al., 2013; Wasan et al., 2014). As shown in Figure 2.1, all data collection, DNA extraction, targeted sequencing and genotyping was completed by others prior to the commencement of this project (Smith et al., 2013; Wasan et al., 2014). A detailed description of the assay methods for somatic tumour DNA samples can be found in Chapter 2.2.1. In brief, somatic mutations were genotyped using a combination of two mutation detection platforms; pyrosequencing and Sequenom. MSI status was determined using the markers BAT-25 and BAT-26 (Smith et al., 2013). The data resulting from this previous work were obtained by the candidate as a Microsoft Excel file of clinical molecular data (Table 2.2). Analyses of the data contained within this file are the focus of this chapter.

3.2.2 Study design and statistical analysis methods

All analyses in this chapter were conducted retrospectively. No stratification by disease stage was employed (due to all patients having Stage IV disease). The primary endpoint was OS; the time from trial randomisation to death.

3.2.2.1 Inter- and intra-genic mutation correlations

Inter- and intra-genic mutations were analysed in R. The Chi-square Test, or Fisher's Exact Test where appropriate ($n < 5$), were calculated to analyse mutation co-occurrences using the chisq.test and fisher.test functions from the base stats package. The recode function from the car package was used to recode mutation information for each gene into binary format in order to group mutations into respective codons and an overall mutant group. The cor function from the base stats package was used to compute the correlations between all mutations. The rcorr function from Hmisc was used to create a matrix of these correlations. The correlation plot describing the inter- and intra-locus correlations between *KRAS*, *BRAF* and *NRAS* mutations and MSI status was created using the corrplot package. The correlation plot colours were chosen using the brewer.pal function from the RColorBrewer package. Microsoft PowerPoint was used to add additional annotations to the plot, which included additional axes labels and the overlay of black boxes to highlight the correlations that remained significant after correction for multiple testing. The Bonferroni correction for multiple testing was performed for 480 independent tests between inter- and inter-genic mutations. Statistically significant findings therefore had to be of the magnitude $P < 1.0 \times 10^{-4}$ in order to retain statistical significance.

3.2.2.2 Clinicopathological analyses

Clinicopathological analyses were analysed in R. The chisq.test and fisher.test functions from the base stats package were used to perform the Chi-square Test, or Fisher's Exact Test where appropriate ($n < 5$) to assess differences in clinicopathology between patient groups, including differences in primary tumour site and sites of metastases. *KRAS* mutations were analysed on an *NRAS* and *BRAF* wild type background, *BRAF* mutations analysed on a *RAS* wild type and MSS background, *NRAS* mutations analysed on a *KRAS* and *BRAF* wild type background and MSI status was analysed on a *RAS* and *BRAF* wild type background. The Bonferroni correction for multiple testing was performed for 24 independent tests for analyses of *KRAS*, *BRAF* and *NRAS* and eight independent tests for analyses of MSI. Statistically significant findings therefore had to be of the magnitude $P < 2.1\text{x}10^{-3}$ and $P < 6.3\text{x}10^{-3}$, respectively, in order to retain statistical significance.

3.2.2.3 Survival analyses

All survival analyses were performed using Cox proportional hazards regression in R. The coxph function from the survival package was used to perform all regression analyses. The cox.zph function from the survival package was used to test the assumption of proportionality in the coxph model, which held in all cases. The Surv, survfit and survdiff functions from the survival package were used to create a survival object, create survival curves and test differences between survival curves, respectively.

Cox proportional hazards regression was chosen as it is a 'semi-parametric' analysis method. Unlike the majority of other survival models, the baseline hazard function is estimated non-parametrically; therefore the survival times are not assumed to follow a particular statistical distribution (Bradburn et al., 2003). The key assumption underlying this model is that the hazard of the event in any group is a constant multiple of the hazard in any other. The Cox model is the most commonly used approach for analysing survival time data in medical research due to its ability to include covariates in the model, thus enabling multivariable analyses of survival data (Clark et al., 2003; Bradburn et al., 2003).

Three sets of survival analyses were performed in this chapter, for which the ggsurvplot function from the survminer package was used to create all survival curves. Firstly, a comparison of survival data between COIN and COIN-B patients was conducted in order to facilitate the combining of these cohorts, which would give subsequent analyses a higher degree of statistical power. The data was split into four different analysis groups in order to ascertain whether any significant differences in survival existed between the COIN and COIN-B datasets. These consisted of (I) COIN vs. COIN-B, (II) each trial arm individually, (III) chemotherapy regimen (OxMdG vs. XELOX) and (IV) cetuximab administration (yes vs. no). Additionally, Cochran's Q Tests were performed and the I^2 statistic calculated for each group using the metagen and summary functions of the meta R package to ensure no significant heterogeneity existed between them before the COIN and COIN-B datasets were combined for all subsequent analyses.

Secondly, survival analyses were performed grouped by somatic mutation and MSI status. *KRAS* mutations were analysed on an *NRAS* and *BRAF* wild type background, *BRAF* mutations analysed on a *RAS* wild type and MSS background, *NRAS* mutations analysed on a *KRAS* and *BRAF* wild type background, and MSI status was analysed on a *RAS* and *BRAF* wild type background in order to avoid the potential confounding effects of known somatic prognostic markers identified in the previous mutation correlation analyses. The recode function from the car package was used to group individual mutations into their respective coding regions to be analysed by codon, and to their respective gene to be analysed as an entire mutant group.

Both univariable and multivariable analyses were performed, the latter including the following prognostic covariates available from the clinical molecular dataset (Table 2.2); age at randomisation, sex, WHO PS, resection status of primary tumour, site of primary tumour, WBC count, ALKP, platelet count, number of metastatic sites, site of distant metastases, cetuximab administration (yes or no), chemotherapy schedule (continuous or intermittent), type of chemotherapy (OxMdG or XELOX), trial status (COIN or COIN-B) and the rs9929218 genotype (homozygous for the minor allele vs. heterozygous/homozygous for the major allele), which have all been shown to have an impact on CRC prognosis (Kohne et al., 2002; Smith et al., 2015; Li, 2016). Tests for interaction were performed within Cox models in order to ascertain whether mutation status was prognostic or predictive for cetuximab treatment, chemotherapy regimen and chemotherapy schedule.

Thirdly, survival analyses were performed with groups split by whether patients had received cetuximab in order to identify whether this influenced survival. Additionally, Cochran's Q and I^2 tests of heterogeneity were performed for these groups to ensure no significant heterogeneity existed between them.

The Bonferroni correction for multiple testing was performed for 30 independent tests for survival analyses split by somatic mutations and MSI. Statistically significant findings therefore had to be of the magnitude $P < 1.6\text{x}10^{-3}$ in order to retain statistical significance.

3.2.2.4 Power analyses

Statistical power calculations were performed for each group using the online calculator at http://www.sample-size.net/sample-size-survival-analysis/. These data were tabulated using Microsoft Excel and power curves were produced using the ggplot2 package in R.

3.3 Results

3.3.1 Frequency of somatic *KRAS*, *BRAF* and *NRAS* mutant and MSI tumour samples

Work undertaken by others prior to the commencement of this project sought to screen for somatic mutations in tumour samples collected from 2671 patients in the MRC clinical trials COIN and COIN-B, as previously described (Smith et al., 2013; Wasan et al., 2014). Fourteen mutations in *KRAS* were screened for, and *KRAS* mutations were identified in a total of 858 patients (39.9% [858/2152] of all patients with *KRAS* data). Two mutations in *BRAF* were screened for, and *BRAF* mutations were identified in a total of 199 patients (9.5% [199/2097] of all patients with *BRAF* data). Nine mutations in *NRAS* were screened for, and *NRAS* mutations were identified in a total of 83 patients (4.0% [83/2091] of all patients with *NRAS* data). The presence of MSI was identified in a total of 66 patients (present in 4.2% [66/1567] of all patients with MSI status data) (Table 3.1).

Table 3.1: Frequency of somatic mutations and MSI-positive tumours in COIN & COIN-B.

Gene/event	Mutation/codon	Frequency (%)
KRAS	Any *KRAS* mutation	858/2152 (39.9)
	c.34G>A (p.G12S)	46/858 (5.4)
	c.34G>C (p.G12R)	14/858 (1.6)
	c.34G>T (p.G12C)	76/858 (8.9)
	c.35G>A (p.G12D)	251/858 (29.3)*
	c.35G>C (p.G12A)	55/858 (6.4)
	c.35G>T (p.G12V)	215/858 (25.0)
	c.37G>A (p.G13S)	1/858 (0.1)
	c.37G>C (p.G13R)	1/858 (0.1)
	c.37G>T (p.G13C)	8/858 (0.9)*
	c.38G>A (p.G13D)	157/858 (18.3)*
	c.38G>T (p.G13V)	1/858 (0.1)
	c.182A>G (p.Q61R)	8/858 (0.9)
	c.182A>T (p.Q61L)	9/858 (1.0)
	c.183A>C (p.Q61H)	22/858 (2.6)*
BRAF	Any *BRAF* mutation	199/2097 (9.5)
	c.1781A>G (p.D594G)	21/199 (10.6)
	c.1799T>A (p.V600E)	178/199 (89.4)
NRAS	Any *NRAS* mutation	83/2091 (4.0)
	c.34G>T (p.G12C)	18/83 (21.7)
	c.35G>A (p.G12D)	2/83 (2.4)
	c.35G>T (p.G12V)	1/83 (1.2)
	c.37G>C (p.G13R)	1/83 (1.2)
	c.38G>A (p.G13D)	1/83 (1.2)
	c.181C>A (p.Q61K)	29/83 (34.9)
	c.182A>G (p.Q61R)	18/83 (21.7)
	c.182A>T (p.Q61L)	12/83 (14.5)
	c.183A>C (p.Q61H)	1/83 (1.2)
MSI	MSI	66/1567 (4.2)

**KRAS* mutations total 100.6% due to multiple mutations in some patients; G12D and G13C: 1, G12D and G13D: 2, G12D and Q61H: 1, G13D and Q61H: 1. Of the 2671 samples, 514 failed *KRAS* genotyping and five were uncharacterised mutants (not analysed). 2152 samples were analysed, of which 858 were *KRAS* mutant. 574 failed *BRAF* genotyping, leaving 2097 samples, of which 199 were *BRAF* mutant. 579 failed *NRAS* genotyping and one was an uncharacterised mutant, leaving 2091 tumour samples, of which 83 were *NRAS* mutant.1104 failed genotyping, leaving 1567 samples, of which 66 were MSI-positive.

3.3.2 Inter- and intra-genic mutation correlations

Cross-correlations between somatic mutations were determined to facilitate future survival stratification. When analysed both individually and by codon, all *KRAS* mutations showed similar effects in terms of mutual exclusivity. Codon 12 (4 of 627 mutant CRCs), 13 (4 of 161) and 61 (2 of 35) mutations were rarely found together. Only *BRAF* c.1799T$>$A (p.V600E) and specific mutations in *NRAS* codon 61 shared this characteristic. Only 1.1% (2/178) of *BRAF* c.1799T$>$A (p.V600E) mutant CRCs had *RAS* mutations compared to 46.9% (894/1908) of *BRAF* wild type CRCs ($P < 2.2\text{x}10^{-16}$, $P < 1.1\text{x}10^{-13}$ after the Bonferroni correction for multiple testing was applied). In contrast, more c.1781A$>$G (p.D594G) mutations co-occurred with *RAS* mutations (14.3% [3/21]) as compared to *BRAF* c.1799T$>$A (p.V600E; P=9.0x10^{-3}); albeit less commonly than found in *BRAF* wild type CRCs (P=3.0x10^{-3}). One case of c.34G$>$A (p.G12S) was noted which co-occurred with *BRAF* c.1799T$>$A (p.V600E; P=2.5x10^{-3} as compared to other *KRAS* mutations [1/812 co-occurred]). For *NRAS*, only 5.0% (3/60) of codon 61 mutant CRCs had *KRAS* mutations compared to 43.5% (10/23) of codons 12 and 13 mutant CRCs (P=7.9x10^{-5}, P=0.04 after correction); the latter being at a similar level to that found in *NRAS* wild type CRCs (40.0% [808/2018], P=0.98). Differences were also observed in the relationship between *BRAF* mutations and MSI status. *BRAF* c.1799T$>$A (p.V600E) was strongly associated with MSI (11.2% [20/178] *BRAF* c.1799T$>$A (p.V600E) CRCs had MSI compared to 2.4% [46/1,908] *BRAF* wild type CRCs, P=5.3x10^{-10}, P=2.5x10^{-7} after correction), whereas c.1781A$>$G (p.D594G) and MSI did not co-occur (0/21). Observations that were significantly positively and negatively correlated are shown in Figure 3.1.

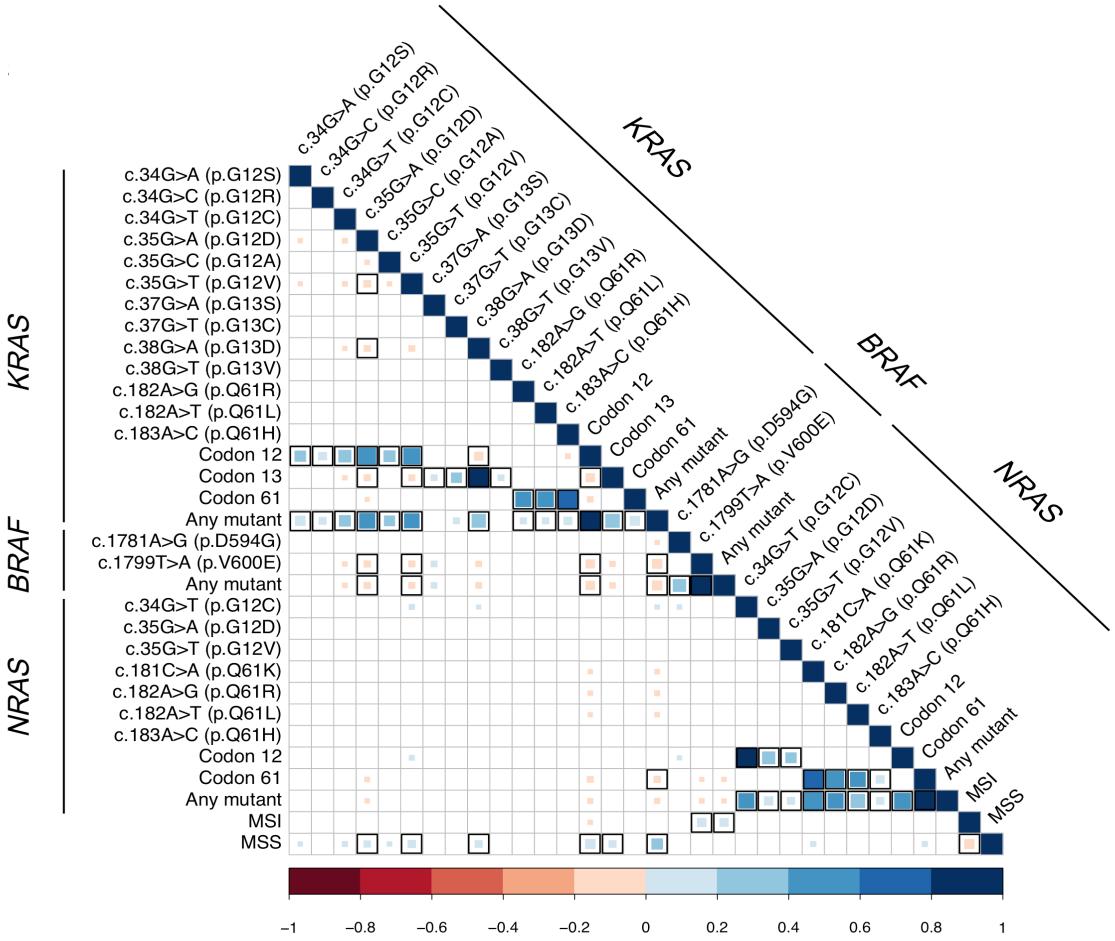

Figure 3.1: Correlations between somatic mutations and MSI status in mCRC. Blue: Positive correlation. Red: Negative correlation. Significant correlations are shown ($P < 0.05$), with the size of the boxes corresponding to the size of the correlation; those that remained significant after correction for multiple testing are outlined in black. *KRAS* codon 12 and 13 mutations rarely co-occurred with *BRAF* V600E mutations: 1/660 *KRAS* codon 12 mutations and 1/178 *BRAF* V600E mutations co-occurred, P=1.4x10^{-13} and 1/168 *KRAS* codon 13 mutations and 1/178 *BRAF* V600E mutations co-occurred, P=2.2x10^{-3}. *KRAS* codon 12 mutations and *NRAS* codon 61 mutations rarely co-occur: 2/660 *KRAS* codon 12 mutations and 2/60 *NRAS* codon 61 mutations co-occurred, P=3.1x10^{-4}. MSI co-occured with *BRAF* V600E: 20/178 (11.2%) *BRAF* V600E mutations and 20/66 (30.3%) with MSI co-occurred, P=2.2x10^{-16} (this was not the case for *BRAF* D594G). *BRAF* D594G co-occured with *RAS* mutations more often than V600E: 3/21 (14.3%) *RAS* mutations on a *BRAF* D594G background and 2/178 (1.1%) *RAS* mutations on a *BRAF* V600E background, P=0.009. *NRAS* codon 12 13 mutations co-occured with *KRAS* mutations more often than codon 61 mutations: 10/23 (43.5%) *KRAS* mutations on a *NRAS* codon 12 and 13 background and 3/60 (5.0%) *KRAS* mutations on an *NRAS* codon 61 background, P=7.9x10^{-5}.

3.3.3 Clinicopathological analyses

3.3.3.1 *KRAS*

More *KRAS* mutant tumours were found in the right colon (257 patients with tumours in the right colon who had *KRAS* mutant CRC out of 436 with *KRAS* mutant or *KRAS* wild type CRC [58.9%]) as compared to the left colon (526 patients with tumours in the left colon who had *KRAS* mutant CRC out of 1331 with *KRAS* mutant or *KRAS* wild type CRC [39.5%], P=2.0x10^{-12}) and more were associated with lung metastases (358 patients with lung metastases who had *KRAS* mutant CRC out of 715 with *KRAS* mutant or *KRAS* wild type CRC [50.1%]) as compared to liver only (156 patients with metastases in the liver only who had *KRAS* mutant CRC out of 418 with *KRAS* mutant or *KRAS* wild type CRC [37.3%], P=4.2x10^{-5}) (Table 3.2).

In terms of codon-specific mutations, more *KRAS* codon 12 and 13 mutant tumours were found in the right colon (32.6% [248/760] versus 17.9% [179/1002], P=1.2x10^{-12}), less in the left colon (66.3% [504/760] versus 80.3% [805/1002], P=3.7x10^{-11}) and more were associated with metastases in the lung (45.0% [342/760] versus 35.6% [357/1002], P=8.4x10^{-5}) and less in liver only (20.0% [152/760] versus 26.1% [262/1002], P=3.1x10^{-3}), as compared to *KRAS* wild type tumours; the correlations for right colon, left colon and lung remained significant after correction for multiple testing (Table 3.2). More *KRAS* codon 61 mutant patients had peritoneal metastases (27.3% [9/33] versus 13.3% [133/1002], P=0.04) as compared to *KRAS* wild type patients. However, there were no significant differences in clinicopathology between *KRAS* codons 12 and 13 versus *KRAS* codon 61 mutant patients (Table 3.2).

Table 3.2: Clinicopathology according to *KRAS* mutation status.

Characteristics		Mutation frequency[1]	Codon 12 & 13 (n=760)	Codon 61 (n=33)	Wild type (n=1002)	*P* (Codon 12 & 13 vs. Wild type)	*P* (Codon 61 vs. Wild type)	*P* (Codon 12 & 13 vs. Codon 61)
Sex	Female	289/593 (48.7)	275 (36.2)	14 (42.4)	304 (30.3)	0.01	0.20	0.59
	Male	503/1201 (41.9)	485 (63.8)	19 (57.6)	698 (69.7)	0.01	0.20	0.59
Age	Mean	NA	63	61	63	NA	NA	NA
Primary tumour [2,3]	Right colon	257/436 (58.9)	248 (32.6)	10 (30.3)	179 (17.9)	$1.2\text{x}10^{-12}$ [$2.9\text{x}10^{-11}$]	0.11	1.00
	Left colon	526/1331 (39.5)	504 (66.3)	22 (66.6)	805 (80.3)	$3.7\text{x}10^{-11}$ [$8.9\text{x}10^{-10}$]	0.09	0.92
Metastatic site [3]	Liver only	156/418 (37.3)	152 (20.0)	4 (12.1)	262 (26.1)	$3.1\text{x}10^{-3}$	0.07	0.37
	Liver	598/1356 (44.1)	577 (75.9)	22 (66.6)	758 (75.6)	0.94	0.33	0.32
	Nodal	359/832 (43.1)	345 (45.4)	15 (45.5)	473 (47.2)	0.48	0.98	1.00
	Lung	358/715 (50.1)	342 (45.0)	16 (48.5)	357 (35.6)	$8.4\text{x}10^{-5}$ [$2.0\text{x}10^{-3}$]	0.18	0.83
	Peritoneum	126/259 (48.6)	117 (15.4)	9 (27.3)	133 (13.3)	0.23	0.04	0.11

KRAS mutations were analysed on an *NRAS* and *BRAF* wild type background (n=1794; total number of mutations=1795 as one patient had multiple mutations). Of these, 760 were found within *KRAS* codon 12 & 13, 33 within *KRAS* codon 61, and 1002 were *KRAS* wild type tumours. [1]There was a significant difference between *KRAS* mutant CRCs in the primary tumour location ($P=2.0\text{x}10^{-12}$) and in sites of metastases ($P=4.6\text{x}10^{-4}$) compared to wild type CRCs. [2]Right colon includes right colon, transverse colon and caecum. Left colon includes left colon, sigmoid colon, rectosigmoid junction and rectum. Percentages are shown in regular parentheses. [3]Some patients had multiple mutations or missing primary tumour site data and some patients had multiple metastases, therefore not all percentages equal 100%. P-values remaining significant after multiple testing correction are shown in square parentheses. NA: Not applicable.

3.3.3.2 *BRAF*

More *BRAF* mutant tumours were found in the right colon (56 patients with tumours in the right colon who had *BRAF* mutant CRC out of 172 with *BRAF* mutant or *BRAF* wild type CRC [32.6%]) as compared to the left colon (57 patients with tumours in the left colon who had *BRAF* mutant CRC out of 618 with *BRAF* mutant or *BRAF* wild type CRC [9.2%], P=2.8x10^{-14}), and more were associated with peritoneal metastases (25 patients with peritoneal metastases who had *BRAF* mutant CRC out of 107 with *BRAF* mutant or *BRAF* wild type CRC [23.4%]) as compared to liver only (22 patients with metastases in the liver only who had *BRAF* mutant CRC out of 214 with *BRAF* mutant or *BRAF* wild type CRC [10.3%], P=3.1x10^{-3}; Table 3.3).

In terms of individual mutations, *BRAF* c.1781A$>$G (p.D594G) tumours had similar clinicopathology to *BRAF* wild type tumours. In contrast, more *BRAF* c.1799T$>$A (p.V600E) tumours were found in the right colon (56.0% [56/100] versus 17.5% [121/693], $P<$2.2x10^{-16}) and more were associated with peritoneal metastases (24.0% [24/100] versus 11.8% [82/693], P=1.5x10^{-3}) as compared to *BRAF* wild type tumours; these correlations remained significant after correction for multiple testing (Table 3.3).

In terms of intra-locus differences, there was a significant difference between c.1781A$>$G (p.D594G) and c.1799T$>$A (p.V600E) with respect to the location of the primary tumour due to fewer c.1781A$>$G (p.D594G) tumours in the right colon (6.6% [1/15] versus 56.0% [56/100], P=1.0x10^{-3}, P=0.02 after correction). There was no significant difference between the sites of metastases associated with these mutations (Table 3.3).

Table 3.3: Clinicopathology according to *BRAF* mutation status.

Characteristics		Mutation frequency[1]	D594G (n=15)	V600E (n=100)	Wild type (n=693)	*P* (D594G vs. Wild type)	*P* (V600E vs. Wild type)	*P* (D594G vs. V600E)
Sex	Female	55/249 (22.1)	7 (46.6)	48 (48.0)	194 (28.0)	0.20	8.0×10^{-5} [1.9×10^{-3}]	1.00
	Male	60/559 (10.1)	8 (53.3)	52 (52.0)	499 (72.0)	0.20	8.0×10^{-5} [1.9×10^{-3}]	1.00
Age	Mean	NA	67	63	63	NA	NA	NA
Primary tumour [2,3]	Right colon	56/172 (32.6)	1 (6.6)	56 (56.0)	121 (17.5)	0.49	$<2.2 \times 10^{-16}$	1.0×10^{-3} [0.02]
	Left colon	57/618 (9.2)	14 (93.3)	44 (44.0)	571 (82.4)	0.49	$<2.2 \times 10^{-16}$	1.0×10^{-3} [0.02]
Metastatic site [3]	Liver only	22/214 (10.3)	3 (20.0)	19 (19.0)	192 (27.7)	0.77	0.09	1.00
	Liver	83/619 (13.4)	13 (86.6)	70 (70.0)	536 (77.3)	0.54	0.14	0.23
	Nodal	53/368 (14.4)	7 (46.6)	46 (46.0)	315 (45.5)	1.00	1.00	1.00
	Lung	35/272 (12.9)	6 (40.0)	29 (29.0)	237 (34.2)	0.85	0.36	0.57
	Peritoneum	25/107 (23.4)	1 (6.6)	24 (24.0)	82 (11.8)	1.00	1.5×10^{-3} [0.04]	0.19

BRAF mutations were analysed on a *RAS* wild type and MSS background (n=808). Of these, 15 were found at *BRAF* D594G, 100 at *BRAF* V600E, and 693 were *BRAF* wild type tumours. [1]There was a significant difference between *BRAF* mutant CRCs in primary tumour location ($P=2.8 \times 10^{-14}$) and metastasic sites ($P=0.03$) compared to wild type CRCs. [2]Definitions of right and left colon as per Table 3.2. Percentages are shown in regular parentheses. [3]Some patients had multiple mutations or missing primary tumour site data and some patients had multiple metastases, therefore not all percentages equal 100%. P-values remaining significant after multiple testing correction are shown in square parentheses. NA: Not applicable. D594G: c.1781A>G; p.D594G. V600E: c.1799T>A; p.V600E.

3.3.3.3 *NRAS*

There was no difference between the frequency of *NRAS* mutant and wild type CRCs with respect to the location of the primary tumour (Table 3.4). However, more *NRAS* mutant CRCs were associated with metastases in the lung (43 patients with lung metastases who had *NRAS* mutant CRC out of 400 with *NRAS* mutant or *NRAS* wild type CRC [10.8%]) as compared to liver only (10 patients with metastases in the liver only who had *NRAS* mutant CRC out of 272 with *NRAS* mutant or *NRAS* wild type CRC [3.7%], P=1.4x10^{-3}).

In terms of individual codons, *NRAS* codon 12 and 13 mutant tumours showed similar clinicopathology to *NRAS* wild type tumours. *NRAS* codon 61 mutant CRCs had similar primary tumour distributions but significantly fewer liver only (12.3% [7/57] versus 26.1% [262/1002], P=0.03) and more lung metastases (68.4% [39/57] versus 35.6% [357/1002], P=1.3x10^{-6}, P=3.1x10^{-5} after correction) as compared to *NRAS* wild type tumours. There were no significant differences in clinicopathology between *NRAS* codons 12 and 13 versus *NRAS* codon 61 mutant CRCs (Table 3.4).

Table 3.4: Clinicopathology according to *NRAS* mutation status.

Characteristics		Mutation frequency[1]	Codon 12 & 13 (n=11)	Codon 61 (n=57)	Wild type (n=1002)	*P* (Codon 12 & 13 vs. Wild type)	*P* (Codon 61 vs. Wild type)	*P* (Codon 12 & 13 vs. Codon 61)
Sex	Female	20/324 (6.2)	2 (18.2)	18 (31.6)	304 (30.3)	0.52	0.96	0.49
	Male	48/746 (6.4)	9 (81.8)	39 (68.4)	698 (69.7)	0.52	0.96	0.49
Age	Mean	NA	59	62	63	NA	NA	NA
Primary tumour [2,3]	Right colon	12/191 (6.3)	2 (18.2)	10 (17.5)	179 (17.9)	1.00	1.00	1.00
	Left colon	52/857 (6.1)	9 (81.8)	43 (75.4)	805 (80.3)	1.00	0.47	1.00
Metastatic site [3]	Liver only	10/272 (3.7)	3 (27.3)	7 (12.3)	262 (26.1)	1.00	0.03	0.35
	Liver	52/810 (6.4)	8 (72.7)	44 (77.2)	758 (75.6)	0.74	0.92	0.71
	Nodal	35/508 (6.9)	3 (27.3)	32 (56.1)	473 (47.2)	0.23	0.24	0.11
	Lung	43/400 (10.8)	4 (36.4)	39 (68.4)	357 (35.6)	1.00	$1.3x10^{-6}$ [$3.1x10^{-5}$]	0.08
	Peritoneum	5/138 (3.6)	1 (9.1)	4 (7.0)	133 (13.3)	1.00	0.22	1.00

NRAS mutations were analysed on a *KRAS* and *BRAF* wild type background (n=1070). Of these, 11 were found within *NRAS* codon 12 & 13, 57 within *NRAS* codon 61, and 1002 were *NRAS* wild type tumours. [1]There was a significant difference between *NRAS* mutant CRCs in the sites of metastases ($P=2.5x10^{-3}$) as compared to wild type CRCs. [2]Definitions of right and left colon as per Table 3.2. Percentages are shown in regular parentheses. [3]Some patients had multiple mutations or missing primary tumour site data and some patients had multiple metastases, therefore not all percentages equal 100%. P-values remaining significant after multiple testing correction are shown in square parentheses. NA: Not applicable.

3.3.3.4 MSI

More MSI-positive tumours were found in the right colon (14 patients with tumours in the right colon who had MSI-positive CRC out of 130 with MSI-positive or MSS CRC [10.8%]) as compared to the left colon (15 patients with tumours in the left colon who had MSI-positive CRC out of 578 with MSI-positive or MSS CRC [2.6%], P=6.2x10^{-5}, P=5.0x10^{-4} after correction) and less were associated with liver metastases (14 patients with liver metastases out of 29 who had MSI-positive CRC [48%] versus 536 patients with liver metastases out of 693 who had MSS CRC [77%], P=7.3x10^{-4}, P=5.8x10^{-3} after correction) as compared to MSS CRCs (Table 3.5).

There was a significant difference between MSI-positive and MSS CRCs regarding the location of the primary tumour due to more MSI-positive tumours in the right colon (48.3% [14/29] versus 16.7% [116/693], P=4.4x10^{-5}, P=3.5x10^{-4} after correction) and fewer in the left colon (51.7% [15/29] versus 81.0% [561/693], P=3.1x10^{-4}, P=2.5x10^{-3} after correction; Table 3.5).

Table 3.5: Clinicopathology according to MSI status.

Characteristics		**MSI frequency[1]**	**MSI (n=29)**	**MSS (n=693)**	***P* (MSI vs. MSS)**
Sex	Female	11/205 (5.4)	11 (37.9)	194 (28.0)	0.34
	Male	18/517 (3.5)	18 (62.1)	499 (72.0)	0.34
Age	Mean	NA	58	63	NA
Primary tumour [2,3]	Right colon	14/130 (10.8)	14 (48.3)	116 (16.7)	4.4x10^{-5} [3.5x10^{-4}]
	Left colon	15/578 (2.6)	15 (51.7)	561 (81.0)	3.1x10^{-4} [2.5x10^{-3}]
Metastatic sites [3]	Liver only	3/195 (1.5)	3 (10.3)	192 (27.7)	0.05
	Liver	14/550 (2.5)	14 (48.3)	536 (77.3)	7.3x10^{-4} [5.8x10^{-3}]
	Nodal	17/332 (5.1)	17 (58.6)	315 (45.5)	0.23
	Lung	6/243 (2.5)	6 (20.7)	237 (34.2)	0.19
	Peritoneum	7/89 (7.9)	7 (24.1)	82 (11.8)	0.09

MSI status was analysed on a *RAS* and *BRAF* wild type background (n=722). Of these, 29 were MSI-positive and 693 were MSS tumours. [1]There was a significant difference between MSI CRCs in primary tumour location (P=6.2x10^{-5}) as well as in the sites of metastases (P=0.02) compared to MSS CRCs. [2]Right colon includes right colon, transverse colon and caecum. Left colon includes left colon, sigmoid colon, rectosigmoid junction and rectum. Percentages are shown in regular parentheses. [3]Some patients had multiple mutations or missing primary tumour site data and some patients had multiple metastases, therefore not all percentages equal 100%. P-values remaining significant after multiple testing correction are shown in square parentheses. MSI: Microsatellite instability. MSS: Microsatellite stable. NA: Not applicable.

3.3.4 Survival analyses

3.3.4.1 Comparison of COIN and COIN-B

Survival analyses of COIN and COIN-B found no significant difference in OS for the following patient categories; (I) COIN vs. COIN-B (HR 1.06, 95% CI 0.90-1.24, P=0.49; Figure 3.2A), (II) individual trial arms (P=0.41; Figure 3.2B), (III) chemotherapy regimen (OxMdG vs. XELOX; HR 0.98, 95% CI 0.89-1.07, P=0.60; Figure 3.3A), (IV) cetuximab administered (yes vs. no; HR 0.96, 95% CI 0.88-1.05, P=0.41; Figure 3.3B).

I^2 and Cochran's Q tests found no significant heterogeneity between any of these groups (Table 3.6), therefore the COIN and COIN-B datasets were combined in all further survival analyses to achieve greater statistical power.

Table 3.6: Heterogeneity tests for COIN & COIN-B survival analyses.

Patient group analysed	I^2 (%)	Q	P^{HET}
COIN vs. COIN-B	0.00	0.94	0.33
Individual treatment arms	0.00	2.40	0.49
Chemotherapy regimen (OxMdG vs. XELOX)	0.00	0.55	0.46
Cetuximab administered (yes vs. no)	26.1	1.35	0.24

OxMdG: Oxaliplatin modified De Gramont. I^2: I^2 Test of heterogeneity. Q: Cochran's Q value. P^{HET}: Cochran's Q-Test P-value.

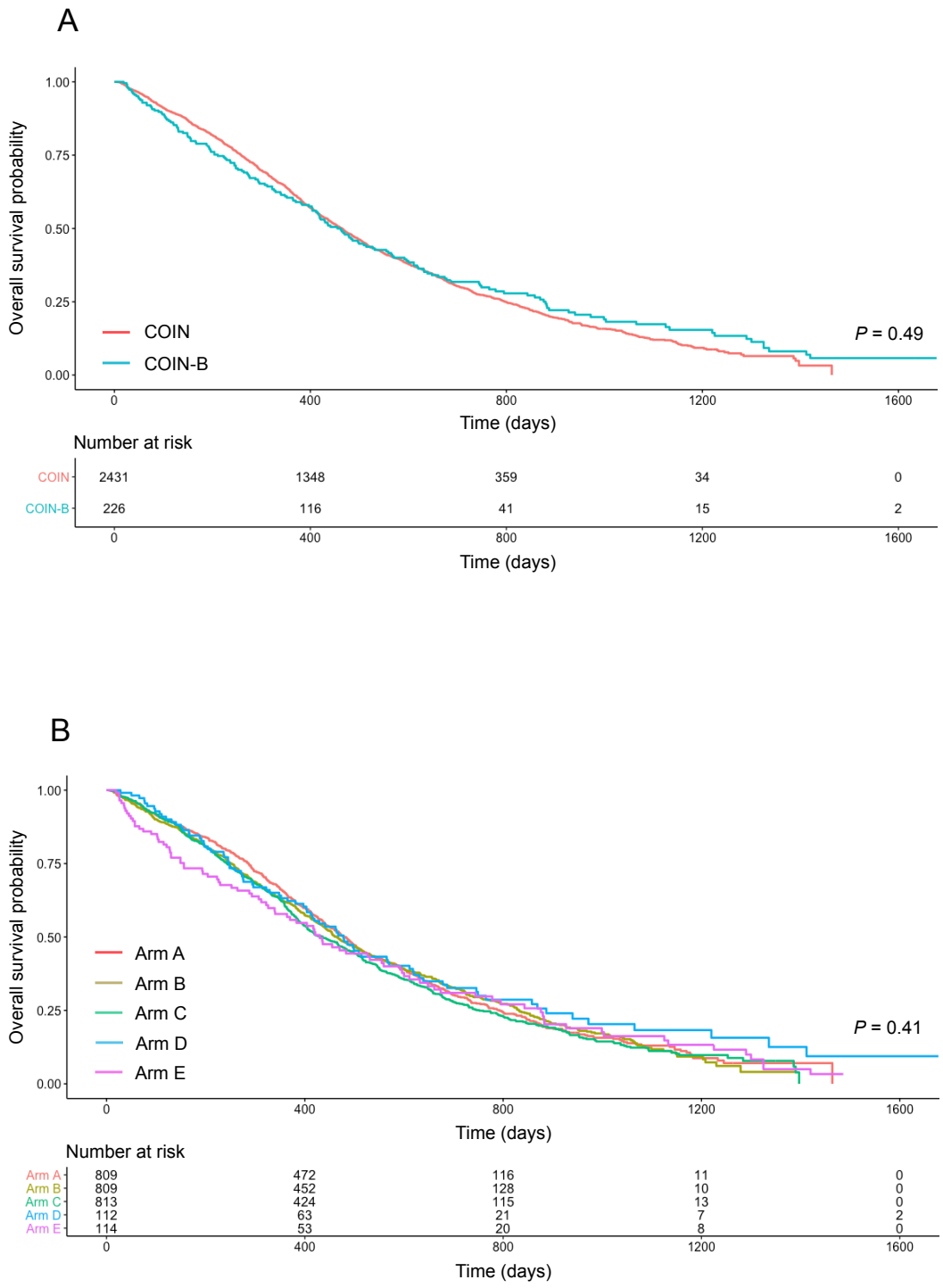

Figure 3.2: OS for patients in COIN & COIN-B by trial status. A: COIN vs. COIN-B. B: Individual trial arms (Arm A: Continuous chemotherapy, Arm B: Continuous chemotherapy plus cetuximab, Arm C: Intermittent chemotherapy, Arm D: Intermittent chemotherapy plus intermittent cetuximab, Arm E: Intermittent chemotherapy plus continuous cetuximab). *P*: Cox Proportional Hazards regression P-value

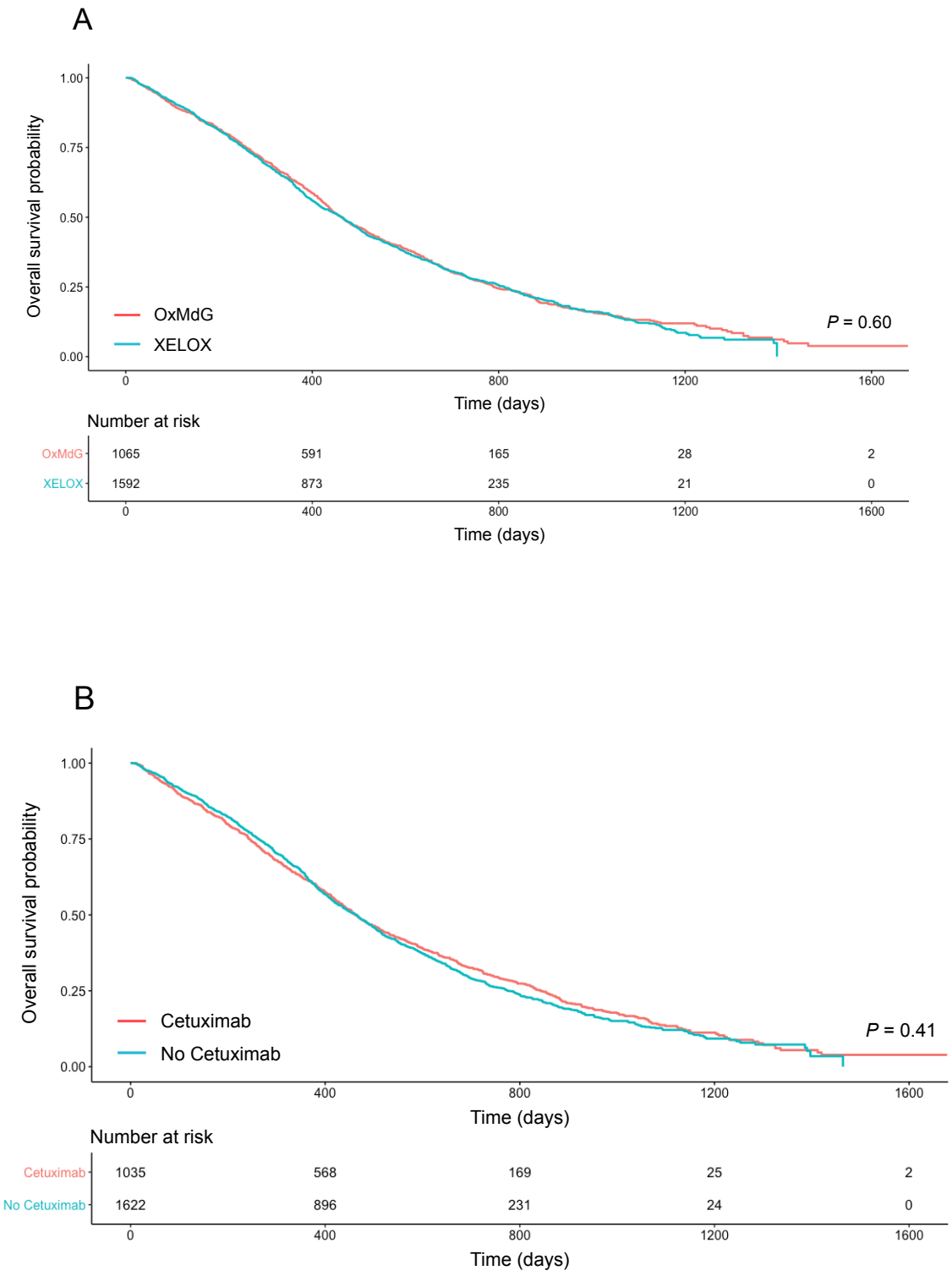

Figure 3.3: OS for patients in COIN & COIN-B by treatment status. A: OxMdG vs. XELOX. B: Cetuximab administration (yes vs. no). OxMdG: Oxaliplatin modified de Gramont. *P*: Cox Proportional Hazards regression P-value

3.3.4.2 Somatic mutations and MSI

The effects of individual, and groups of, somatic mutations on survival were investigated under both univariable and multivariable models (taking into account the previously observed cross-correlations). Results for OS under both models can be found in Table 3.8.

3.3.4.2.1 Power calculations

Survival analyses had $\geq$80% power to detect associations with OS for variants with HRs $\geq$ 1.17 for the *KRAS* mutant cohort (1322 events), variants with HRs $\geq$ 1.37 for the *BRAF* mutant cohort (575 events), variants with HRs $\geq$ 1.50 for the *NRAS* mutant cohort (743 events) and variants with HRs $\geq$ 1.78 for the MSI cohort (499 events), using a two-sided significance level of α=0.05 (Figure 3.4).

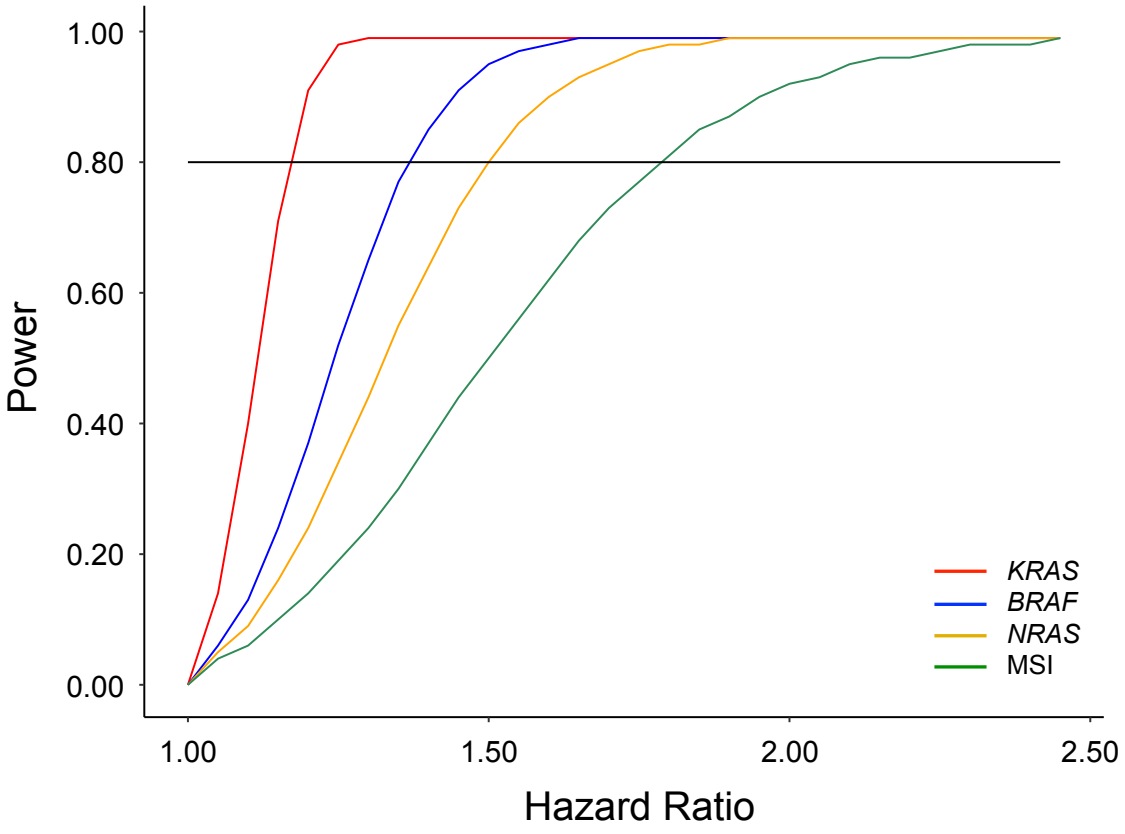

Figure 3.4: Statistical power to detect associations with OS. Statistical power for associations with overall survival for patients with mutations in *KRAS* on a *BRAF* and *NRAS* wild type background (1322 events, red line), *BRAF* on a *RAS* wild type and MSS background (575 events, blue line), *NRAS* on a *KRAS* and *BRAF* wild type background (743 events, orange line) and by MSI status on a *RAS* and *BRAF* wild type background (499 events, green line). Horizontal line represents the threshold for 80% power. MSI: Microsatellite instability. MSS: Microsatellite stable.

3.3.4.2.2 Univariable analyses

KRAS

KRAS mutations conferred a poor prognosis (HR 1.45, 95% CI 1.30-1.61, P=1.9x10^{-11}, median reduction in survival of 131 days). When grouped by codons, both codon 12 and 13 mutations conferred poor prognosis (HR 1.44, 95% CI 1.28-1.61, P=6.4 x10^{-10}, P=1.9x10^{-8} after correction, and HR 1.53, 95% CI 1.26-1.86, P=1.5x10^{-5}, P=4.5x10^{-4} after correction, respectively), whereas codon 61 mutations did not (HR 1.23, 95% CI 0.84-1.81, P=0.28) (Figure 3.5 [red line: codon 12 mutant, green line: codon 13 mutant, blue line: codon 61 mutant, purple line: wild type]); these intra-locus differences were not significant. Five *KRAS* mutations (c.34G > A [p.G12S], c.35 G > A [p.G12D], c.35G > C [p.G12A], c.35 G > T [p.G12V] and c.38G > A [p.G13D]) individually showed significantly poorer prognosis with a median reduction in survival of 213, 111, 65, 160 and 165 days, respectively; four of these remained significant after correction for multiple testing (Table 3.8).

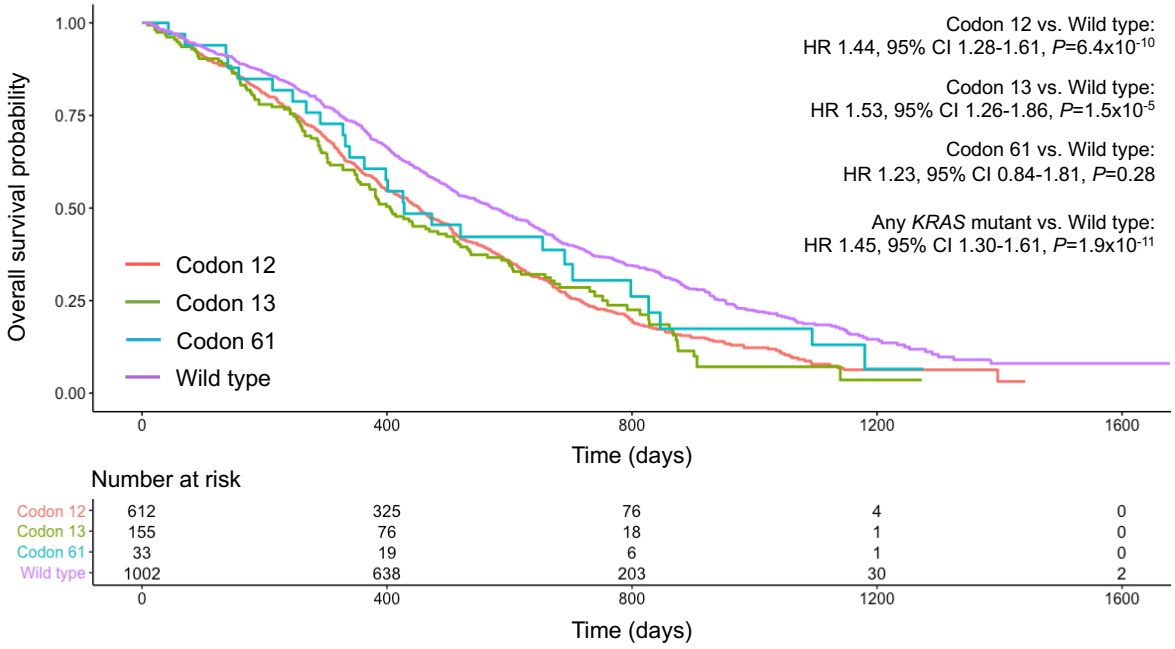

Figure 3.5: OS for patients in COIN & COIN-B by *KRAS* status. HR: Hazard ratio. CI: Confidence interval. *P*: Cox Proportional Hazards regression P-value.

BRAF

BRAF mutations conferred a poor prognosis (HR 2.31, 95% CI 1.85-2.87, P=7.8x10^{-14}, median reduction in survival of 295 days). Patients with *BRAF* c.1799T>A (p.V600E) mutant tumours had significantly poorer prognoses compared to patients with *BRAF* wild type tumours (HR 2.60, 95% CI 2.06-3.28, P=1.0x10^{-15}, P=3.0x10^{-14} after correction, median reduction in survival of 320 days) (Figure 3.6 [red line: c.1781A>G (p.D594G) mutant, green line: c.1799T>A (p.V600E) mutant, blue line: wild type]), whereas there was no evidence to suggest that patients with c.1781A>G (p.D594G) mutant tumours had a significantly different prognosis to patients with *BRAF* wild type tumours (HR 1.30, 95% CI 0.73-2.31, P=0.37; Table 3.8).

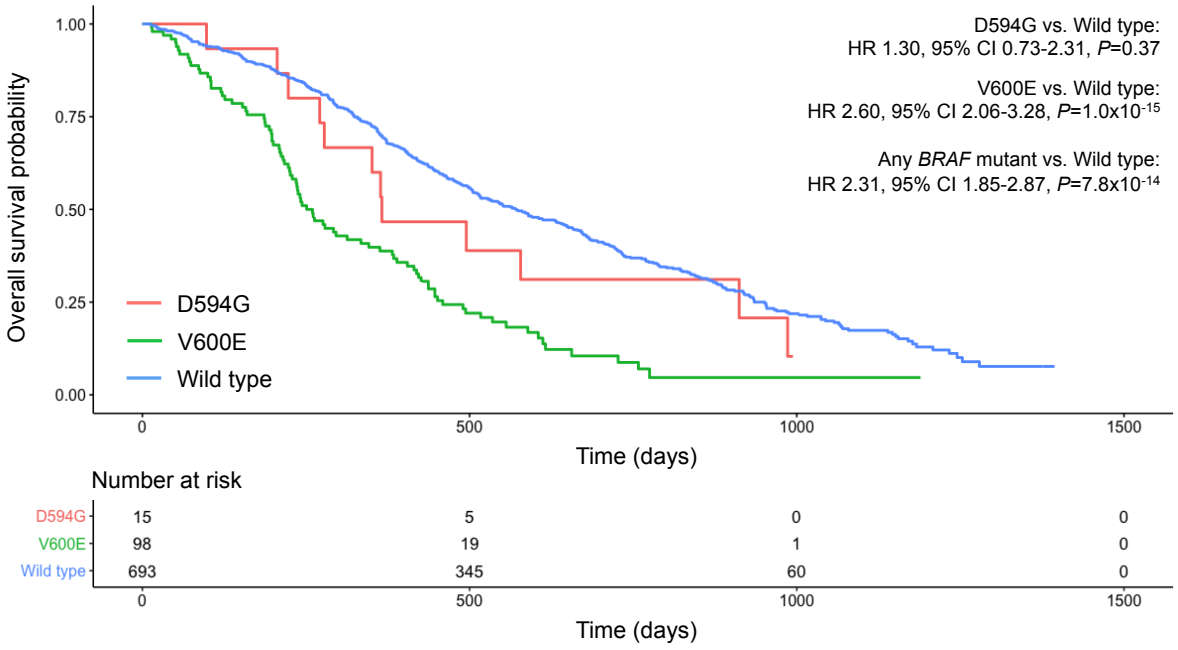

Figure 3.6: OS for patients in COIN & COIN-B by *BRAF* status. D594G: c.1781A>G (p.D594G). V600E: c.1799T>A (p.V600E). HR: Hazard ratio. CI: Confidence interval. P: Cox Proportional Hazards regression P-value.

NRAS

NRAS mutations conferred a poor prognosis (HR 1.44, 95% CI 1.09-1.90, *P*=0.01, median reduction in survival of 112 days). When grouped by codon, patients with codon 61 mutant tumours conferred a significantly poorer prognosis to patients with *NRAS* wild type tumours (HR 1.47, 95% CI 1.09-1.99, *P*=0.01, median reduction in survival of 131 days) (Figure 3.7 [red line: codons 12 and 13 mutant, green line: codon 61 mutant, blue line: wild type]), whereas there was no evidence to suggest that patients with codon 12 or 13 mutant tumours had a significantly different prognosis to patients with *BRAF* wild type tumours (HR 1.29, 95% CI 0.64-2.58, *P*=0.48; Table 3.8). However, when analysed individually, *NRAS* mutations showed no significant differences in survival.

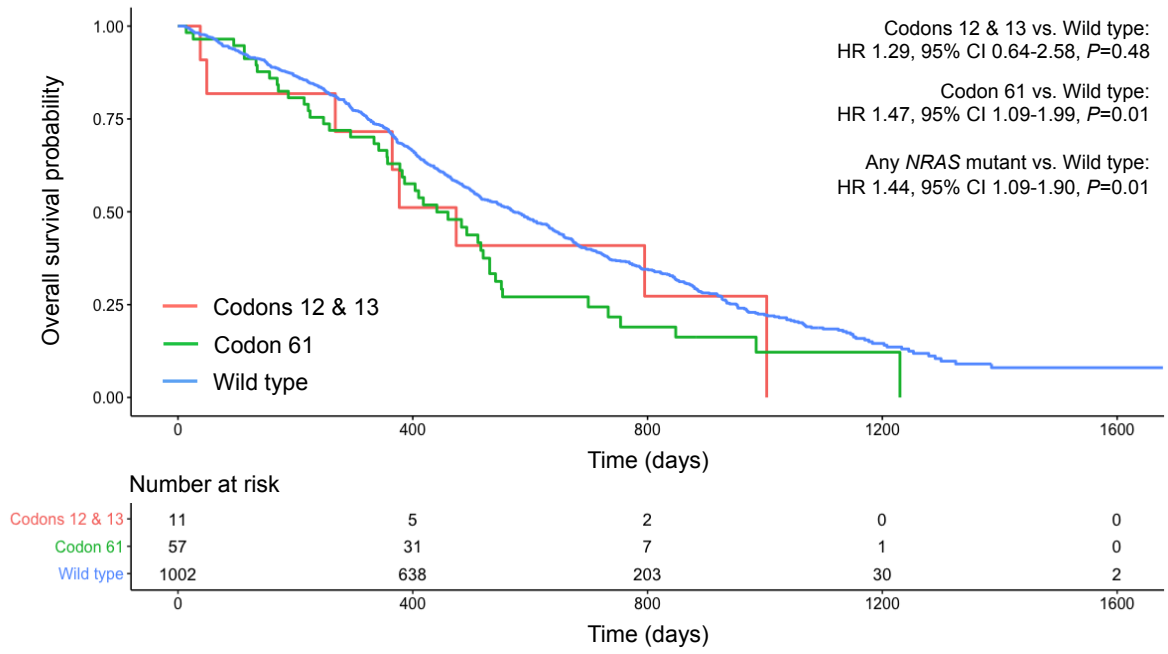

Figure 3.7: OS for patients in COIN & COIN-B by *NRAS* status. HR: Hazard ratio. CI: Confidence interval. *P*: Cox Proportional Hazards regression P-value.

MSI

Patients with MSI-positive CRC had a significantly inferior prognosis compared to those with MSS tumours (HR 1.86, 95% CI 1.22-2.83, P=4.0x10^{-3}, median reduction in survival of 244 days; Table 3.8) (Figure 3.8 [red line: MSI, blue line: MSS]), although this association did not withstand the Bonferroni correction for multiple testing.

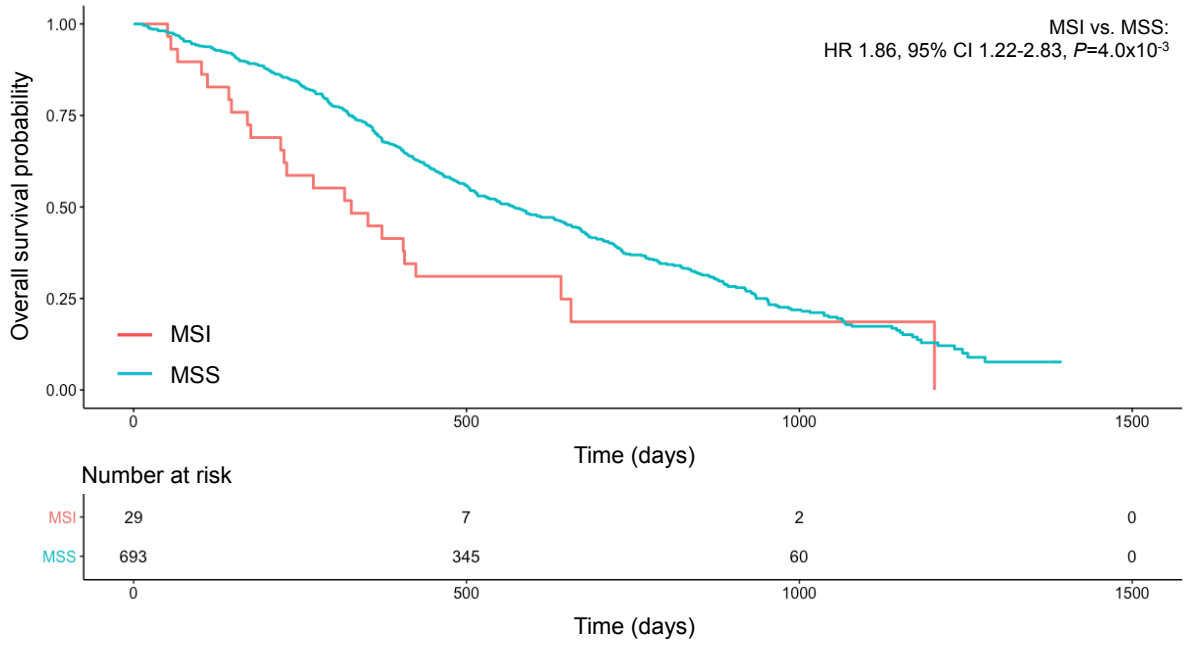

Figure 3.8: OS for patients in COIN & COIN-B by MSI status. MSI: Microsatellite instability. MSS: Microsatellite stable. HR: Hazard ratio. CI: Confidence interval. *P*: Cox Proportional Hazards regression P-value.

3.3.4.2.3 Multivariable analyses

Following the inclusion of known prognostic factors into the model (Table 3.7), similar statistically significant associations between somatic variants and OS were identified for patients with *KRAS* c.34G > A (p.G12S), c.35G > A (p.G12D), c.35G > T (p.G12V) and c.38G > A (p.G13D) mutant tumours, codon 12 and codon 13 mutant tumours and for all *KRAS* mutations combined compared *KRAS* wild type tumours. All of these associations withstood correction for multiple testing (Table 3.8).

Similar associations between somatic variants and prognosis were also identified for patients with *BRAF* c.1799T > A (p.V600E) and for all *BRAF* mutations combined compared with *BRAF* wild type tumours. All of these associations withstood correction for multiple testing (Table 3.8). The significant difference in OS between patients with *BRAF* c.1781A > G (p.D594G) and *BRAF* c.1799T > A (p.V600E) mutant tumours previously identified through univariable analyses was no longer significant after correction for prognostic factors (HR 0.97, 95% CI 0.41-2.31, *P*=0.94).

Patients with *NRAS* c.181C > A (p.Q61K) mutant tumours had significantly poorer prognosis compared to patients with *NRAS* wild type tumours when analysed in multivariable analyses (HR 1.63, 95% CI 1.03-2.60, *P*=0.04) whereas univariable analyses did not identify a significant difference in OS between the two patient groups (HR 1.43, 95% CI 0.96-2.21, *P*=0.11). However, this significant difference in OS did not withstand correction for multiple testing. Multivariable analyses also increased the significance of two prognostic associations previously identified through univariable analyses; *NRAS* codon 61 mutations versus wild type (univariable HR 1.47, 95% CI 1.09-1.99, *P*=0.01, multivariable HR 1.69, 95% CI 1.22.2.34, $P=1.8\text{x}10^{-3}$) and all *NRAS* mutations combined versus wild type (univariable HR 1.44, 95% CI 1.09-1.90, *P*=0.01, multivariable HR 1.56, 95% CI 1.16-2.11, $P=3.7\text{x}10^{-3}$; Table 3.8). However, neither of these associations withstood correction for multiple testing.

A similar outcome of poor prognosis for patients with MSI tumours through univariable analyses (HR 1.86, 95% CI 1.22-2.83, $P=4.0\text{x}10^{-3}$) was also identified through multivariable analyses, although it was no longer found to be statistically significant (HR 1.36, 95% CI 0.78-2.35, *P*=0.27; Table 3.8). Interaction terms for chemotherapy regimen, schedule and cetuximab administration with *KRAS*, *BRAF*, *NRAS* and MSI status were also added into the model, but none of these were significant.

Table 3.7: Individual results for covariates in the multivariable analysis model.

Covariate	**HR**	**95% CI**	***P***
Age	1.00	0.99-1.00	0.56
Sex (male)	0.87	0.78-0.97	9.7×10^{-3}
WHO PS	1.42	1.31-1.56	$< 2.0 \times 10^{-16}$
Primary tumour resection status*	1.29	1.01-1.63	0.04
Primary tumour site (right colon)$^{+}$	1.36	1.16-1.60	1.5×10^{-4}
WBC count	1.03	1.03-1.04	$< 2.0 \times 10^{-16}$
ALKP level	1.00	1.00-1.00	$< 2.0 \times 10^{-16}$
PLT count	1.00	1.00-1.00	$< 2.0 \times 10^{-16}$
Number of metastatic sites	1.21	1.14-1.28	2.5×10^{-10}
Liver metastases	1.23	1.09-1.39	8.8×10^{-4}
Lung metastases	0.98	1.02-0.88	0.66
Peritoneal metastases	1.34	1.16-1.54	6.4×10^{-5}
Nodal metastases	1.15	1.04-1.28	7.5×10^{-3}
Other metastases	1.31	1.15-1.51	7.9×10^{-5}
Trial (COIN)	1.10	0.92-1.33	0.30
Chemotherapy regimen (XELOX)	1.03	0.93-1.15	0.57
Chemotherapy schedule (intermittent)	1.09	0.98-1.21	0.10
Cetuximab administration (yes)	0.95	0.85-1.05	0.31
rs9929218 genotype (AA)	1.41	1.19-1.69	1.1×10^{-4}

*: Unresected/unresectable vs. local recurrence. $^{+}$: Association of primary tumour site remained significant in stratified *KRAS*, *BRAF*, *NRAS* and MSI survival analyses. HR: Hazard ratio. CI: Confidence interval. *P*: P-value. WHO: World Health Organization. PS: Performance status. WBC: White blood cell. ALKP: Alkaline phosphatase. PLT: Platelet. Nodal metastases were included in the model but are excluded here due to NA values for all somatic mutations and MSI. NA: Not applicable.

Table 3.8: OS for patients in COIN & COIN-B stratified by somatic mutation and MSI status.

			univariable analyses			multivariable analyses		
Mutation/codon	**Cases**	**Events**	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
				KRAS				
c.34G>A (p.G12S)	42	35	1.78	1.27-2.50	$9.2\text{x}10^{-4}$ (0.03)	1.82	1.26-2.64	$1.5\text{x}10^{-3}$ (0.05)
c.34G>C (p.G12R)	14	10	0.95	0.51-1.78	0.88	0.84	0.42-1.66	0.61
c.34G>T (p.G12C)	69	53	1.21	0.91-1.60	0.18	1.20	0.87-1.65	0.26
c.35G>A (p.G12D)	236	190	1.48	1.26-1.74	$1.6\text{x}10^{-6}$ ($4.8\text{x}10^{-5}$)	1.36	1.13-1.65	$1.4\text{x}10^{-3}$ (0.04)
c.35G>C (p.G12A)	51	41	1.43	1.04-1.96	0.03	1.42	1.01-1.99	0.04
c.35G>T (p.G12V)	200	162	1.48	1.25-1.76	$7.5\text{x}10^{-6}$ ($2.3\text{x}10^{-4}$)	1.54	1.26-1.87	$1.8\text{x}10^{-5}$ ($5.4\text{x}10^{-4}$)
c.37G>A (p.G13S)	-	-	-	-	-	-	-	-
c.37G>C (p.G13R)	-	-	-	-	-	-	-	-
c.37G>T (p.G13C)	8	6	1.36	0.61-3.03	0.46	0.66	0.21-2.12	0.49
c.38G>A (p.G13D)	146	116	1.53	1.26-1.87	$2.2\text{x}10^{-5}$ ($6.6\text{x}10^{-4}$)	1.55	1.23-1.94	$1.6\text{x}10^{-4}$ ($4.8\text{x}10^{-3}$)
c.38G>T (p.G13V)	1	1	-	-	-	-	-	-
c.182A>G (p.Q61R)	8	6	1.41	0.63-3.15	0.41	2.27	0.92-5.60	0.07
c.182A>T (p.Q61L)	7	6	1.27	0.57-2.84	0.56	1.40	0.62-3.16	0.42
c.183A>C (p.Q61H)	18	15	1.17	0.70-1.95	0.56	1.29	0.72-2.33	0.39
Codon 12 mutant	612	491	1.44	1.28-1.61	$6.4\text{x}10^{-10}$ ($1.9\text{x}10^{-8}$)	1.44	1.26-1.65	$1.7\text{x}10^{-7}$ ($5.1\text{x}10^{-6}$)
Codon 13 mutant	155	123	1.53	1.26-1.86	$1.5\text{x}10^{-5}$ ($4.5\text{x}10^{-4}$)	1.51	1.21-1.89	$3.0\text{x}10^{-4}$ ($9.0\text{x}10^{-3}$)
Codon 61 mutant	33	27	1.23	0.84-1.81	0.28	1.47	0.96-2.25	0.08
Any *KRAS* mutant	800	641	1.45	1.30-1.61	$1.9\text{x}10^{-11}$ ($5.7\text{x}10^{-10}$)	1.46	1.29-1.66	$4.6\text{x}10^{-9}$ ($1.4\text{x}10^{-7}$)
KRAS wild type	1002	691	1.00	Ref.	Ref.	1.00	Ref.	Ref.
				BRAF				

c.1781A>G (p.D594G)	15	12	1.30	0.73-2.31	0.37	1,52	0.79-2.91	0.21
c.1799T>A (p.V600E)	98	87	2.60	2.06-3.28	$1.0x10^{-15}$ ($3.0x10^{-14}$)	2.40	1.82-3.17	$5.3x10^{-10}$ ($1.6x10^{-8}$)
Any *BRAF* mutant	113	99	2.31	1.85-2.87	$7.8x10^{-14}$ ($2.3x10^{-12}$)	2.27	1.75-2.94	$6.4x10^{-10}$ ($1.9x10^{-8}$)
BRAF wild type	693	477	1.00	Ref.	Ref.	1.00	Ref.	Ref.
NRAS								
c.34G>T (p.G12C)	6	5	1.42	0.59-3.43	0.43	0.98	0.39-2.42	0.96
c.35G>A (p.G12D)	2	2	-	-	-	-	-	-
c.35G>T (p.G12V)	1	1	-	-	-	-	-	-
c.37G>C (p.G13R)	1	0	-	-	-	-	-	-
c.38G>A (p.G13D)	1	0	-	-	-	-	-	-
c.181C>A (p.Q61K)	26	21	1.43	0.96-2.21	0.11	1.63	1.03-2.60	0.04
c.182A>G (p.Q61R)	18	13	1.58	0.91-2.73	0.11	1.71	0.95-3.11	0.08
c.182A>T (p.Q61L)	12	11	1.51	0.83-2.73	0.18	1.78	0.93-3.41	0.08
c.183A>C (p.Q61H)	1	0	-	-	-	-	-	-
Codons 12 & 13 mutant	11	8	1.29	0.64-2.58	0.48	1.15	0.55-2.38	0.71
Codon 61 mutant	57	45	1.47	1.09-1.99	0.01	1.69	1.22-2.34	$1.8x10^{-3}$
Any *NRAS* mutant	68	53	1.44	1.09-1.90	0.01	1.56	1.16-2.11	$3.7x10^{-3}$
NRAS wild type	1002	691	1.00	Ref.	Ref.	1.00	Ref.	Ref.
MSI								
MSI	29	23	1.86	1.22-2.83	$4.0x10^{-3}$	1.36	0.78-2.35	0.27
MSS	693	477	1.00	Ref.	Ref.	1.00	Ref.	Ref.

KRAS mutants (versus wild type) were analysed on a *BRAF* and *NRAS* wild type background; *BRAF* mutants (versus wild type) were analysed on a *RAS* wild type and MSS background; *NRAS* mutants (versus wild type) were analysed on a *KRAS* and *BRAF* wild type background; and MSI (versus MSS) was analysed on a *RAS* and *BRAF* wild type background. HRs, CIs and P-values not shown for mutations where number of events < 2. HR: Hazard Ratio. CI: Confidence Interval. Ref.: Referent.

3.3.4.3 Somatic mutations and MSI by cetuximab administration

Survival analyses of somatic mutations and MSI status were split by cetuximab administration. There were no significant differences in OS between patients who did and did not receive cetuximab, in agreement with previous reports (Maughan et al., 2011). Results of survival analyses split by cetuximab administration are shown in Table 3.9.

Table 3.9: OS for patients in COIN & COIN-B stratified by cetuximab administration.

	No Cetuximab administered					Cetuximab administered					Heterogeneity tests		
Mutation/codon	**N**	**E**	**HR**	**95% CI**	***P***	**N**	**E**	**HR**	**95% CI**	***P***	**I^2 %**	**Q**	***P^{HET}***
KRAS													
c.34G>A (p.G12S)	23	20	1.77	1.13-2.78	0.01	19	15	1.80	1.07-3.03	0.03	0.00	0.00	0.97
c.34G>C (p.G12R)	9	7	1.14	0.51-2.57	0.75	5	3	0.61	0.20-1.92	0.40	0.00	0.76	0.38
c.34G>T (p.G12C)	54	42	1.22	0.89-1.68	0.22	15	11	1.09	0.60-1.99	0.78	0.00	0.11	0.74
c.35G>A (p.G12D)	145	116	1.37	1.12-1.69	$2.5x10^{-3}$	91	74	1.68	1.30-2.18	$8.7x10^{-5}$	30.1	1.43	0.23
c.35G>C (p.G12A)	40	31	1.27	0.88-1.84	0.19	11	10	2.02	1.07-3.79	0.03	33.7	1.51	0.22
c.35G>T (p.G12V)	119	96	1.36	1.09-1.70	$6.7x10^{-3}$	81	66	1.68	1.28-2.21	$1.6x10^{-4}$	30.5	1.44	0.23
c.37G>A (p.G13S)	-	-	-	-	-	-	-	-	-	-	-	-	-
c.37G>C (p.G13R)	-	-	-	-	-	-	-	-	-	-	-	-	-
c.37G>T (p.G13C)	6	6	2.59	1.16-5.82	0.02	2	0	-	-	-	-	-	-
c.38G>A (p.G13D)	96	75	1.49	1.17-1.91	$1.4x10^{-3}$	50	41	1.59	1.14-2.21	$6.2x10^{-3}$	0.00	0.08	0.78
c.38G>T (p.G13V)	-	-	-	-	-	1	1	-	-	-	-	-	-
c.182A>G (p.Q61R)	5	4	2.33	0.87-6.25	0.09	3	2	0.84	0.21-3.40	0.81	26.4	1.36	0.24
c.182A>T (p.Q61L)	4	3	0.79	0.25-2.46	0.69	3	3	2.83	0.90-8.87	0.07	58.4	2.41	0.12
c.183A>C (p.Q61H)	10	10	1.31	0.70-2.46	0.40	8	5	0.95	0.39-2.31	0.91	0.00	0.34	0.53
Codon 12 mutant	390	312	1.35	1.17-1.56	$5.9x10^{-5}$	222	179	1.60	1.32-1.94	$1.4x10^{-6}$	48.8	1.95	0.16
Codon 13 mutant	102	81	1.54	1.22-1.96	$4.0x10^{-4}$	53	42	1.49	1.08-2.07	0.02	0	0.03	0.87
Codon 61 mutant	19	17	1.30	0.80-2.11	0.29	14	10	1.15	0.61-2.16	0.67	0.00	0.09	0.78
Any *KRAS* mutant	511	410	1.38	1.21-1.58	$3.4x10^{-6}$	289	231	1.55	1.30-1.85	$1.2x10^{-6}$	2.70	1.03	0.31
KRAS wild type	598	423	1.00	Ref.	Ref.	404	268	1.00	Ref.	Ref.	-	-	-
BRAF													

c.1781A>G (p.D594G)	11	10	1.44	0.77-2.70	0.26	4	2	0.85	0.21-3.45	0.82	0.00	0.45	0.35
c.1799T>A (p.V600E)	75	66	2.49	1.90-3.26	$3.1x10^{-11}$	23	21	2.77	1.74-4.42	$1.9x10^{-5}$	0.00	0.16	0.69
Any *BRAF* mutant	86	76	2.26	1.76-2.91	$2.2x10^{-10}$	27	23	2.31	1.48-3.62	$2.4x10^{-4}$	0.00	0.01	0.94
BRAF wild type	478	339	1.00	Ref.	Ref.	215	138	1.00	Ref.	Ref.	-	-	-
NRAS													
c.34G>T (p.G12C)	2	1	0.57	0.08-4.03	0.57	4	4	2.29	0.85-6.16	0.10	35.7	1.56	0.21
c.35G>A (p.G12D)	-	-	-	-	-	2	2	1.16	0.29-4.68	0.83	-	-	-
c.35G>T (p.G12V)	-	-	-	-	-	1	1	-	-	-	-	-	-
c.37G>C (p.G13R)	-	-	-	-	-	1	0	-	-	-	-	-	-
c.38G>A (p.G13D)	-	-	-	-	-	1	0	-	-	-	-	-	-
c.181C>A (p.Q61K)	13	11	1.32	0.72-2.40	0.37	13	10	1.61	0.85-3.03	0.14	0.00	0.20	0.65
c.182A>G (p.Q61R)	11	7	1.25	0.59-2.64	0.56	7	6	2.28	1.01-5.15	0.05	11.3	1.13	0.29
c.182A>T (p.Q61L)	4	4	2.07	0.77-5.56	0.15	8	7	1.36	0.64-2.89	0.42	0.00	0.44	0.13
c.183A>C (p.Q61H)	-	-	-	-	-	1	0	-	-	-	-	-	-
Codons 12 & 13 mutant	2	1	0.57	0.08-4.03	0.57	9	7	1.63	0.77-3.45	0.20	0.00	0.97	0.32
Codon 61 mutant	28	22	1.39	0.90-2.13	0.14	29	23	1.61	1.05-2.46	0.03	0.00	0.23	0.63
Any *NRAS* mutant	30	23	1.30	0.86-1.98	0.22	38	30	1.61	1.10-2.35	0.01	0.00	0.53	0.46
NRAS wild type	598	423	1.00	Ref.	Ref.	404	268	1.00	Ref.	Ref.	-	-	-
MSI													
MSI	14	12	2.01	1.13-3.58	0.02	15	11	1.83	0.98-3.39	0.06	0.00	0.05	0.82
MSS	478	339	1.00	Ref.	Ref.	215	138	1.00	Ref.	Ref.	-	-	-

KRAS mutants (versus wild type) were analysed on a *BRAF* and *NRAS* wild type background; *BRAF* mutants (versus wild type) were analysed on a *RAS* wild type and MSS background; *NRAS* mutants (versus wild type) were analysed on a *KRAS* and *BRAF* wild type background; and MSI (versus MSS) was analysed on a *RAS* and *BRAF* wild type background. HRs, CIs and P-values not shown for mutations where number of events < 2. N: Sample size. E: Events. HR: Hazard Ratio. CI: Confidence Interval. Ref.: Referent. P^{HET}: P=value of heterogeneity.

3.4 Discussion

3.4.1 Distinguishing between driver and passenger mutations through mutational co-occurrences

Although all cancers arise as a result of somatically acquired changes in the DNA of cancer cells, this does not mean that all somatic abnormalities present in a cancer genome have been involved in the cancer's development, with some potentially having had no contribution at all (Stratton et al., 2009). The identification of mutations that drive tumourigenesis is critical for precision oncology (Bailey et al., 2018) and for this reason, the terms 'driver' and 'passenger' have been coined to distinguish between mutations that are causally implicated in tumourigenesis and confer a growth advantage on the cancer cell, and those that do not (Stratton et al., 2009). The analysis of inter- and intra-genic mutational co-occurrences in this chapter highlighted a number of relationships between somatic mutations that give an insight into which mutations are potential drivers of CRC tumourigenesis.

Mutations in codons 12, 13 and 61 of *KRAS* were found to tend towards mutual exclusivity. Considering that mutations affecting these regions of the *KRAS* oncogene normally cause it to become constitutively active and drive cell proliferation (Haigis et al., 2008; Pylayeva-Gupta et al., 2011; Stolze et al., 2015), this finding would indicate that *KRAS* mutations are a likely driver of CRC development. Indeed, this has been demonstrated in human colorectal cancer cell lines (Shirasawa et al., 1993), and somatic *KRAS* mutations are established as early events in colorectal tumourigenesis (Vogelstein et al., 1988; Fearon and Vogelstein, 1990; Walther et al., 2009; Fearon, 2011).

However, different *KRAS* mutants have been shown to be associated with varying biological outcomes (Hobbs et al., 2016a), and is it therefore possible that not all *KRAS* mutations share equal tumourigenic effects (Stinchcombe and Der, 2011; Kirk, 2011; Stolze et al., 2015; Hobbs et al., 2016a; Hobbs and Der, 2019). It has been shown in human isogenic cell lines that expression of three of the most common *KRAS* mutations (G12D, G12V and G13D) are associated with significantly increased proliferation compared with *KRAS* wild type controls (Stolze et al., 2015), which could potentially explain why the few instances of co-occurring *KRAS* mutations identified in the results shown here all include either G12D or G13D point mutations, through the process of subclonal selection (Martincorena et al., 2017). These mutational co-occurrences could potentially be evidence of tumour subclones, although due to the lack of variant allele frequency information in the dataset it is impossible to be certain of the pattern of intratumoral genetic heterogeneity for these patients (Williams et al., 2018).

In terms of *BRAF*, the c.1799T$>$A (p.V600E) mutation is another well established driver of CRC tumourignenesis (Jass, 2007; Walther et al., 2009; Fearon, 2011), which has been shown to be more strongly selected during tumour progression than other *BRAF* mutations (Temko et al., 2018). The results reported here support this, with only 1.1% of *BRAF* c.1799T$>$A

(p.V600E) mutations having co-occurred with *RAS* mutations, in agreement with other reports (Rajagopalan et al., 2002; Hutchins et al., 2011; Imamura et al., 2012; Phipps et al., 2013; Gonsalves et al., 2014; Cremolini et al., 2015). This suggests that *KRAS* and *BRAF* mutations might exhibit equivalent tumourigenic effects, in concurrence with others' findings (Rajagopalan et al., 2002).

In contrast, 14.3% of *BRAF* c.1781A $>$ G (p.D594G) mutations occurred together with *RAS* mutations. The nature of the c.1781A $>$ G (p.D594G) mutation is currently not well understood (Cremolini et al., 2015; Amaki-Takao et al., 2016), although the kinase activity of *BRAF* mutations at codon 594 has previously been reported to be low (Ikenoue et al., 2003; Wan et al., 2004; Smalley et al., 2009). Whilst p.V600E mutations were determined to be activating alterations, p.D594G mutations are assumed to display impaired kinase activity. This supports the differences between *BRAF* c.1781A $>$ G (p.D594G) and *BRAF* c.1799T $>$ A (p.V600E) mutant CRCs reported in the present study. However, considering the frequency with which *BRAF* c.1781A $>$ G (p.D594G) mutations co-occurred with *RAS* mutations was significantly lower than *BRAF* wild type CRCs (14.3% vs. 46.9%, respectively), *BRAF* c.1781A $>$ G (p.D594G) is unlikely to be a benign passenger mutation; it may potentially cause a partial loss of gene function (i.e. it may be hypomorphic) and subsequently play a minor role in colorectal tumourigenesis. Indeed, it has been proposed that some passenger mutations may be better described as 'mini-drivers' (Castro-Giner et al., 2015) due to them displaying similar characteristics as driver mutations, only to a lesser extent (McFarland et al., 2017).

Although mutations in the *NRAS* oncogene are rarer in CRC than those in *KRAS* (Irahara et al., 2010), they have nonetheless been shown to contribute to CRC development and progression (Haigis et al., 2008; Wang et al., 2013). The results of this chapter found that 5.0% of mutations in *NRAS* codon 61 co-occurred with *KRAS* mutations. In contrast, 43.5% of mutations in codons 12 and 13 co-occurred with *KRAS* mutations at a similar level to *NRAS* wild type CRCs. This observation indicates that there might be underlying biological differences between mutations in codon 61 compared with codons 12 and 13 of *NRAS*. Some evidence for the presence of distinct functional effects was reported using murine models for melanoma. Burd et al showed that *Nras* p.Q61R mutants are able to efficiently drive in vivo melanomagenesis, whereas *Nras* p.G12D do not (Burd et al., 2014). Studies have also shown that *NRAS* mutations arise at a later stage in the development of malignancies than *KRAS* mutations (Vogelstein et al., 1988; Demunter et al., 2001), further supporting a potential difference in biological effects in different *NRAS* mutations.

MSI is an indicator of defective MMR, which is critical for the maintenance of genomic stability (Peltomaki, 2003; Fearon, 2011). The involved proteins, such as MLH and MSH, play a 'caretaker' role in tumourigenesis through protecting the genome against mutations (Kinzler and Vogelstein, 1997; van Heemst et al., 2007). The results shown here support previous reports that MMR-deficient tumours (i.e. MSI-positive tumours) have a very high incidence of *BRAF* mutations and a lower incidence of *KRAS* mutations (Rajagopalan et al., 2002). These

findings are consistent with the hypothesis that both *KRAS* and *BRAF* mutant tumours progress through the same biological pathways, but the mutation spectrum depends upon the underlying genomic instability of the tumours (Kinzler and Vogelstein, 1996). Moreover, MSI was found to occur with *BRAF* c.1799T $>$ A (p.V600E) mutations, but not with *BRAF* c.1781A $>$ G (p.D594G) mutations, in agreement with other studies (Tran et al., 2011; Lochhead et al., 2013; Gonsalves et al., 2014; Cremolini et al., 2015; Amaki-Takao et al., 2016), giving further support to their differing biological effects.

3.4.2 The relationship between somatic mutations and clinicopathology

It is important to investigate the relationship between somatic mutations and clinicopathology, as CRC is known to exhibit differences in incidence, pathogenesis, molecular pathways and outcome depending on the location of the primary tumour (Lee et al., 2015), and profound differences in metastatic patterns between histological subtypes and the localisation of the primary tumour in CRC have been observed (Hugen et al., 2014). It has been suggested that through influencing the tumour's biological behaviour, different somatic profiles are associated with different clinicopathology (Tran et al., 2011). This is supported by the results reported here, with a number of mutations conferring differences in the clinicopathology of their primary tumours, as well as their sites of metastases.

In agreement with previous studies (Yamauchi et al., 2012; Rosty et al., 2013), more *KRAS* mutant tumours were observed in the right colon (58.9%) as opposed to the left colon (39.5%; P=2.0x10^{-12}). Differences have been shown to exist between the right (proximal) and left (distal) colon in terms of developmental origin, exposure to patterning genes, environmental mutagens and gut flora (Missiaglia et al., 2014). For instance, the right colon originates from the midgut, whereas the left colon stems from the hindgut (Riihimaki et al., 2016), which may account for the different spectrum of somatic mutations in the left and right colon. Additionally, *KRAS* mutant tumours were most commonly associated with lung metastases (50.1%), which is consistent with a report suggesting the presence of *KRAS* mutations to be a predictor for the development of lung metastases (Pereira et al., 2015).

A higher percentage of *BRAF* mutant tumours were found in the right colon (32.6%) as opposed to the left colon (9.2%), and more *BRAF* mutant tumours were significantly associated with peritoneal metastases compared to *BRAF* wild type tumours (P=1.5x10^{-3}), in agreement with others' findings (Tran et al., 2011; Yokota et al., 2011; Schirripa et al., 2015). A possible reason for this could be that *BRAF* mutant tumours have been shown to exhibit an increased frequency of mucinous differentiation (Pai et al., 2012; Rosty et al., 2013; Schirripa et al., 2015). Metastases have been shown to spread through the peritoneal fluid within the peritoneal cavity, and mucinous adenocarcinomas have been shown to metastasise excessively within the peritoneum (Riihimaki et al., 2016), which has led others to suggest that the production of mucus may enable access to the peritoneal space (Hugen et al., 2014). Mucinous adenocarcinomas

are thought to be more aggressive (Riihimaki et al., 2016), and have been shown to confer a poor prognosis in mCRC (Mekenkamp et al., 2012), which would correlate with the poor prognosis displayed by *BRAF* mutant tumours.

However, clinicopathological differences were observed between *BRAF* c.1781A > G (p.D594G) and *BRAF* c.1799T > A (p.V600E) mutant tumours, with the vast majority of *BRAF* c.1781A > G (p.D594G) mutant tumours located in the left colon (99.3%), at similar levels to *BRAF* wild type tumours (82.4%). This finding is consistent with other studies, which reported that CRCs with *BRAF* c.1781A > G (p.D594G) mutant tumours exhibit similar clinicopathological features to those of *BRAF* wild type tumours (Cremolini et al., 2015; Amaki-Takao et al., 2016).

NRAS mutant tumours were predominantly found in the left colon (81.8% of codon 12 and 13 mutations and 75.4% of codon 61 mutations). This concurs with others' findings (Irahara et al., 2010; Cercek et al., 2017) and adds support to *NRAS* mutations defining a clinically distinct subgroup of mCRC (Cercek et al., 2017). While *KRAS* mutations have been previously associated with right-sided colon tumours, *NRAS* mutations have been associated with tumours in the left colon (Irahara et al., 2010), which suggests a distinct biology for *KRAS* and *NRAS* mutant molecular subsets of CRC (Cercek et al., 2017). However, reports on clinicopathological differences between *KRAS* and *NRAS* mutant CRC have been conflicting, with others' findings indicating that *NRAS* mutant CRCs display similar clinicopathology to *KRAS* mutant CRCs (Schirripa et al., 2015). These differences are potentially due to the rarity of *NRAS* mutations in CRC (Irahara et al., 2010) limiting these studies to a small number of *NRAS* mutant cases (Cercek et al., 2017), which is also a limitation of this study. While no significant differences were identified between *NRAS* mutant and *NRAS* wild type tumours in terms of primary tumour site, significant differences were found with respect to the sites of metastasis. More *NRAS* codon 61 mutant tumours were associated with metastases in the lungs as opposed to *NRAS* wild type tumours (P=1.3x10^{-6}).

In agreement with previous reports (Benatti et al., 2005; Ogino et al., 2009; Tran et al., 2011), tumours harbouring MSI were found in the right colon significantly more often (48.3%) than MSS tumours (16.7%), as expected due to the high co-occurrence of MSI and *BRAF* c.1799T > A (p.V600E) mutations (Benatti et al., 2005). There was a significantly reduced rate of liver metastases between MSI (48.3%) and MSS (77.3%) tumours, which is in agreement with previous reports (Tran et al., 2011). MSI is known to be associated with mucinous histology (Tran et al., 2011), and a previous study has highlighted a reduced rate of liver metastases in mucinous histology mCRC (Catalano et al., 2009), which supports the findings reported here. However, it is unclear as to whether the difference in metastatic spread is due to mucinous histology, MSI, or both (Tran et al., 2011).

3.4.3 The influence of somatic mutations on survival

Previous studies have shown that *KRAS* mutations confer a poor prognosis in CRC patients (Richman et al., 2009; Imamura et al., 2012; Phipps et al., 2013; Eklof et al., 2013), which are supported by the results reported in this thesis. When analysed together, *KRAS* mutant tumours conferred a reduced survival compared to *KRAS* wild type CRCs. This poor prognostic effect was also found when analysed by codon, with *KRAS* codon 12 and 13 mutant tumours conferring a significantly reduced survival rate compared to *KRAS* wild type CRCs.

When analysed together, *BRAF* mutant tumours conferred a poor prognosis in comparison to *BRAF* wild type tumours. Taken individually, tumours with the c.1799T $>$ A (p.V600E) mutation in *BRAF* conferred a significantly worse prognosis to *BRAF* wild type tumours. Patients with *BRAF* c.1781A $>$ G (p.D594G) mutant CRCs also displayed a reduction in OS compared to those with *BRAF* wild type tumours, although this finding was not statistically significant. However, this is likely due to the limited power of the study to detect the prognostic effects of *BRAF* c.1781A $>$ G (p.D594G) mutations, as a result of the small number of patients analysed that harbour this mutation.

Taken as a whole, *NRAS* mutant CRCs conferred an inferior prognosis to that of *NRAS* wild type CRCs, in agreement with another study on the effects of *NRAS* mutations on OS (Schirripa et al., 2015). When analysed individually, *NRAS* codon 61 mutant CRCs conferred a poor prognosis compared to *NRAS* wild type CRCs. Although patients with *NRAS* codon 12 and 13 mutant tumours appeared to have an inferior prognosis to those with *NRAS* wild type tumours, this finding was not significant. This is likely due to the small number of *NRAS* codon 12 and 13 mutations analysed, and a larger sample size would be required in order to investigate this further. There was also no significant difference in survival between *NRAS* codon 61 and *NRAS* codon 12 and 13 mutant tumours. Others' findings are consistent with the findings reported here regarding the prognosis of patients with *NRAS* mutations, claiming that exon 3 (codons 60 and 61) *NRAS* mutants were associated with a strong negative prognostic effect on survival, whereas exon 2 (codons 12 and 13) mutants had no significant difference from *RAS* wild type tumours (Cercek et al., 2017).

The presence of MSI conferred poor prognosis here, in agreement with previous studies of patients with mCRC (Tran et al., 2011; Smith et al., 2013), in contrast to the superior prognosis displayed in patients with locally advanced (Stage II/III) CRC (Popat et al., 2005). Although MSI is known to be associated with *BRAF* mutations, the results here show that MSI has a significant effect on prognosis in mCRC when analysed in patients who have *BRAF* wild type tumours, in agreement with other reports (Tran et al., 2011).

While the observations reported here suggest that somatic mutations in the *KRAS*, *BRAF* and *NRAS* oncogenes and the MSI status of patients' tumours have an effect on mCRC prognosis,

it is important to note that there are also underlying biological reasons for the differences in prognosis between right and left-sided CRCs. Due to the larger bowel lumen, right-sided CRC usually becomes symptomatic later than left-sided CRC, and right-sided tumours are located further away from the anal verge, i.e. they are more difficult to be discovered by digit rectal examination and sigmoidoscopy (Christodoulidis et al., 2010). These factors can lead to patients with right-sided tumours often presenting at a later stage than those with left-sided tumours, which may account for the poorer prognosis associated with these tumours (Christodoulidis et al., 2010).

3.4.4 Conclusion

The results of this chapter show that somatic mutations in the *KRAS*, *BRAF* and *NRAS* oncogenes and MSI status all have a significant effect on CRC prognosis. This informed the decision to perform a genome-wide screen for germline variants associated with CRC prognosis not only on the full patient cohort, but also on a subgroup of patients for which those with somatic mutations and MSI were excluded, in order to unmask the prognostic effects of these mutations. These analyses are the focus of the following chapter.

Chapter 4

The influence of commonly inherited germline variants on survival in mCRC

4.1 Introduction

In addition to the somatic markers discussed in the previous chapter, there is also a need to identify common germline markers that influence CRC prognosis; both in order to extend our current understanding of CRC pathogenesis and to potentially direct clinical management (Phipps et al., 2016). Historically, the search for germline factors associated with survival in CRC has focused predominantly on candidate genes that are known to function either within pathways of action for chemotherapeutic agents used in CRC treatment (Marcuello et al., 2004; Dotor et al., 2006; Walther et al., 2009) or that influence tumour progression (Kim et al., 2008). More recently, a number of studies have focused on variants initially identified as being associated with susceptibility to CRC, and have suggested that some of these alleles may also influence CRC survival (Passarelli et al., 2011; Dai et al., 2012; Phipps et al., 2012; Garcia-Albeniz et al., 2013; Abuli et al., 2013; Takatsuno et al., 2013; Morris et al., 2015).

However, the vast majority of proposed markers have not been validated through independent studies (Tenesa et al., 2010; Hoskins et al., 2012; Sanoff et al., 2014). Independent validation of prognostic biomarker studies is recommended by the REMARK guidelines (McShane et al., 2005) and is essential before variants can be considered for clinical implementation (Erichsen and Chanock, 2004; Van Cutsem et al., 2016). The SNP rs9929218, located in *CDH1*, was found to be associated with susceptibility to CRC through a large meta-analysis of GWAS studies (Houlston et al., 2008) and is currently the only CRC-risk SNP confirmed as being associated with CRC prognosis through a candidate gene study and validated using an independent patient cohort (Smith et al., 2015). The limited robustness of results from previous studies may partially reflect the shortcomings of a candidate-based approach. For instance, the pathways, genes and variants most relevant to and most robustly associated with CRC prognosis may be those without a previously known role in CRC pathogenesis and survival (Phipps et al., 2016). GWASs negate this limitation of candidate gene studies through analysing a large number of SNPs across the whole genome in a relatively unbiased fashion (Cantor et al., 2010).

To date, GWASs have identified 79 germline SNPs significantly associated with CRC susceptibility in European populations (Law et al., 2019) and the GWAS approach may also provide valuable insights into potential prognostic markers for CRC (Phipps et al., 2016). However, while GWAS analyses have identified numerous SNPs robustly associated with the risk of CRC incidence, far fewer markers of survival to CRC have been confirmed through this method. It has been proposed that this is likely to be due to the use of insufficient sample sizes or

the omission of independent validation cohorts with which to corroborate study findings (Smith et al., 2015; Phipps et al., 2016). Currently, the only SNP robustly associated with CRC prognosis through a GWAS is rs209489 at 6p12.1 (Phipps et al., 2016). rs209489 is intronic to *ELOVL5*, which had not previously been implicated in CRC aetiology or progression. Although it is possible that this SNP may reflect the role of another nearby gene, the identification of a gene with no previous association with CRC as having a potential prognostic role highlights the benefits of the GWAS approach (Phipps et al., 2016).

A clear need exists for further study into germline prognostic markers for CRC. Through the use of GWASs with sufficient sample sizes and independent validation cohorts to rigorously test study findings, the opportunity exists to expand upon our current knowledge of the association between germline variants and CRC prognosis, and the potential to gain an improved understanding of CRC pathobiological mechanisms to inform putative clinical applications (Tenesa et al., 2010). The following chapter focuses on GWAS analyses utilising the combined sample sizes of the COIN and COIN-B datasets to identify novel germline SNPs that are associated with CRC survival, and to ascertain whether any support exists within this dataset for both previously proposed and established germline prognostic biomarkers for CRC.

4.1.1 Aims and objectives

- To investigate whether any novel germline mutations are significantly associated with CRC survival through performing GWAS analyses on the combined COIN and COIN-B cohorts
- To perform GWAS analyses on a stratified subset of patients whose tumours are wild type for *RAS* and *BRAF* mutations and are MSS
- To determine whether any support exists for SNPs currently proposed and established as prognostic biomarkers using the COIN and COIN-B patient cohorts

4.2 Materials and methods

4.2.1 Patient and specimen characteristics

Blood DNA samples were available from 1950 cases from the MRC clinical trials COIN (Chapter 2.1.1; n=1780) and COIN-B (Chapter 2.1.2; n=170). Two patients were found to have missing survival data and were excluded, giving a total of 1948 cases to be analysed. As shown in Figure 2.1, all data collection, DNA extraction and genotyping was completed by others prior to the commencement of this project (Al-Tassan et al., 2015). A detailed description of the assay methods for germline DNA samples can be found in Chapter 2.2.2. In brief, initial QC of germline patient data and genotype imputation was performed by Professor Richard Houlston's laboratory at the ICR as previously described (Anderson et al., 2010; Al-Tassan et al., 2015), using data from the 1000 Genomes Project as the haplotype phasing reference panel. The cleaned germline data was obtained by the candidate in the form of 22 Oxford text genotype (.gen) files, the analysis of which is the focus of this chapter.

4.2.2 Study design and statistical analysis methods

All analyses in this chapter were conducted retrospectively. No stratification by disease stage was employed (due to all patients having Stage IV disease). The primary endpoint was OS; the time from trial randomisation to death.

4.2.2.1 Germline data file conversions and preparation

Prior to performing GWAS analyses, the original germline data files for each autosome (chromosomes 1-22) were required to be converted from Oxford text genotype file format (.gen) to transposed .ped (.tped) and .map (.tfam) files, which could be run by the GenABEL software package in R. Firstly, GTOOL was used for .gen to .ped and .map conversion. Due to the size of the .gen files, this file conversion was completed through remote access to the Raven HPC server. Cyberduck was used to transfer files between the local machine and the Raven server. PLINK was then used to convert .ped and .map files into binary .ped and .map (.bed, .bim and .fam) files. Duplicate variant names (such as common occurrences of '.' in place of an rsid identifier) were renamed to a unique identifier using R, through modifying each .bim file. The .bed, modified .bim and .fam files were then converted back into .ped and .map format using PLINK, before being converted to the .tped and .tfam files used for GWAS analyses.

4.2.2.2 MAF filtering

Files were filtered to exclude variants with MAF < 0.05 using PLINK. This choice of MAF threshold was based on statistical power calculations that showed there was a significant increase in power to detect true associations for variants with MAF ≥ 0.05 than those with $0.01 \leq$ MAF < 0.05.

4.2.2.3 Power calculations

Statistical power calculations were performed for the full patient cohort and the all wild type subgroup using the online calculator at http://www.sample-size.net/sample-size-survival-analysis/, tabulated in Microsoft Excel and power curves were produced using the ggplot2 package in R.

4.2.2.4 GWAS analyses

All GWAS analyses were performed using the GenABEL package in R. The convert.snp.tped function was used to convert the separate transposed (.tped and .tfam) files into a single genotype data file for each autosome (chromosomes 1-22). The load.gwaa.data command was used to combine this file with the phenotype information from the clinical molecular dataset (Table 2.2) to create a gwaa.data object ready to be analysed. The GASurv function created a survival data object for use with the mlreg function, which was used for all GWAS analyses. The subset function from the base stats package was used in order to create the all wild type subgroup.

All variants identified as suggestive of association were then analysed under a multivariable Cox survival model in R using the survival package, using the covariates described in Table 3.7. Tests for interaction between these variants and cetuximab use, chemotherapy regimen and chemotherapy schedule were also performed to determine whether any of the variants were prognostic or predictive biomarkers for treatment, although none of these were found to be significant. Following the identification of a genome-wide significant SNP through multivariable analyses of the most highly associated variants in univariable analyses, a full multivariable GWAS was then performed for the full cohort and all wild type subgroup, using the GenABEL package in R through the Raven HPC.

GWAS analyses were performed using an additive model of genetic inheritance. It is common practice for GWASs to examine an additive model only, as this model has sufficient power to detect both additive and dominant effects (Bush and Moore, 2012), although an additive model may be underpowered to detect some recessive effects (Lettre et al., 2007). A recessive model was not considered here due to the multiple testing implications associated with analysing multiple genetic inheritance models (Reed et al., 2015). For two alleles of a biallelic SNP, A and a, the additive model (for allele A) assumes that the increase in risk for each copy of the A allele increases in a uniformly linear fashion (for example, if the effect for genotype Aa is 2x, for AA there will be a 4x increase in effect) (Bush and Moore, 2012).

The genome-wide significance level of $P < 5.0\text{x}10^{-8}$ was originally proposed as an appropriate value by which to achieve a probability of greater than 95% that there are no false positives for 1,000,000 independent association tests (Risch and Merikangas, 1996) and has emerged as the *de facto* standard value for genome-wide significance (Jannot et al., 2015). The threshold of suggestive association is a more arbitrary measure, although SNPs above this threshold are often selected as candidate loci for replication studies (Jin et al., 2013).

Results for each autosome were filtered for genotype imputation fidelity using an inclusion threshold of info score > 0.8 using data from SNPTEST. This threshold was introduced to increase confidence in the results, as SNPs above this threshold are considered to be of high imputation quality (Reed et al., 2015). Results were then combined and visualised graphically using the qqman package in R. The manhattan function was used to create Manhattan plots and the qq function was used to create Q-Q plots. The λ statistic of genomic control was also calculated using the estlambda function in GenAbel in R, in order to be confident that no underlying bias was present due to population stratification. Generally, a value of close to 1.0 is desired to be confident that there is no stratification present. If $\lambda < 1$, no adjustment is required (Hinrichs et al., 2009), while values of $\lambda > 1.2$ suggest the presence of population stratification (Reed et al., 2015).

4.2.2.5 Support for proposed and established germline prognostic biomarkers

Following GWAS analyses, support for previously proposed and established germline prognostic markers was then analysed, comparing survival data from the combined COIN and COIN-B under the respective genetic inheritance models used by previous studies.

4.2.2.6 LD analyses

LD analyses were performed using the ld command in PLINK. All LD values at each locus were calculated with respect to the most significant SNP at that locus (the sentinel SNP).

4.3 Results

4.3.1 Full patient cohort

4.3.1.1 Power calculations

Based on a sample size of 1948 patients (1453 events), GWAS analyses had $\geq$80% power to detect associations with OS for variants with HRs $\geq$ 3.24 and MAFs $\geq$ 0.01 and for variants with HRs$\geq$ 1.71 and MAFs $\geq$ 0.05 (Figure 4.1).

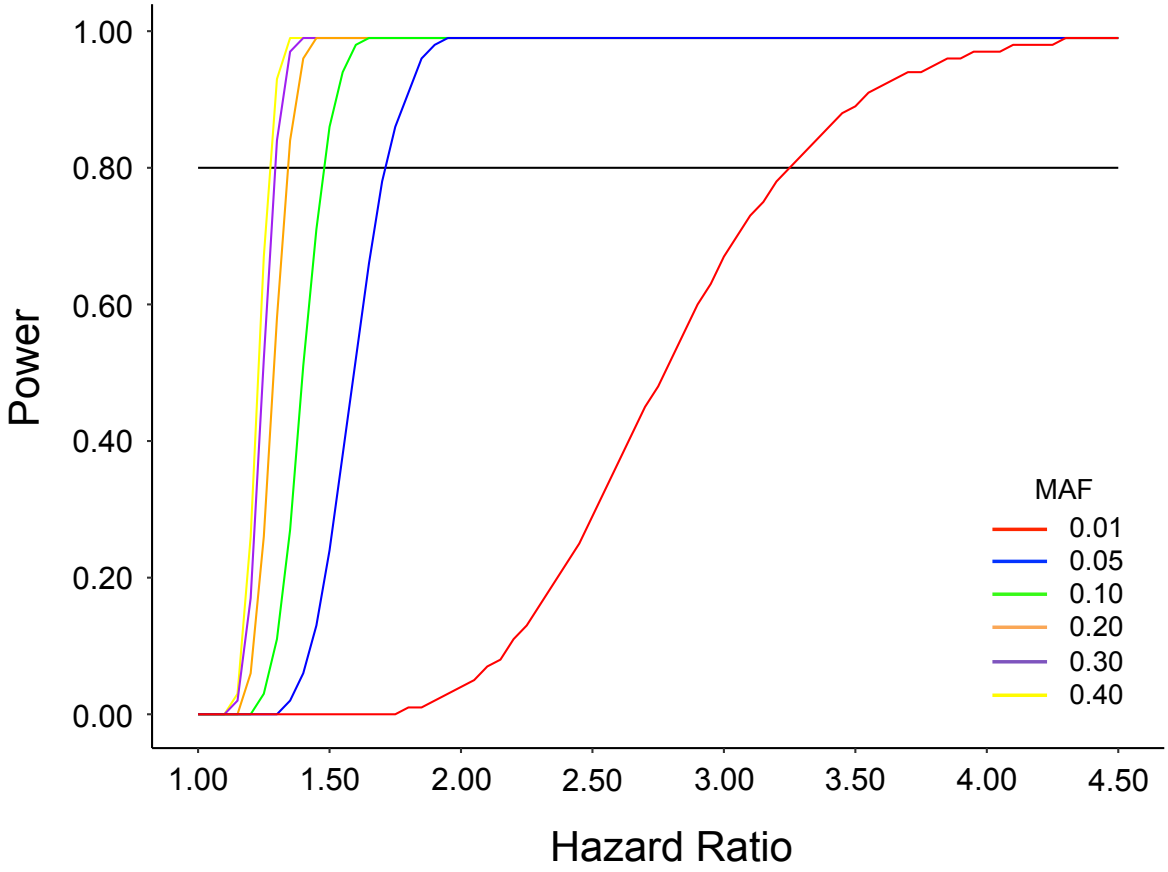

Figure 4.1: Statistical power to detect associations with OS in the full patient cohort. N=1948 (1453 events). α level: $P < 5.0\text{x}10^{-8}$. Horizontal line represents the threshold for 80% power. MAF: Minor allele frequency.

4.3.1.2 GWAS analyses

4.3.1.2.1 Univariable GWAS analyses

Under an additive genetic model for OS, 68 commonly inherited variants (MAF $\geq$ 0.05, info score $>$ 0.8) were found to be suggestive of association with survival ($P < 1.0\text{x}10^{-5}$), including a peak 25 of variants at 1q32.1, of which the most significant SNP was rs706494 (MAF=0.19, info score=0.9; Figure 4.2, Table 4.1). There were no variants associated with survival at the level of genome-wide significance ($P < 5.0\text{x}10^{-8}$) in univariable analyses (Figure 4.2, Table 4.1).

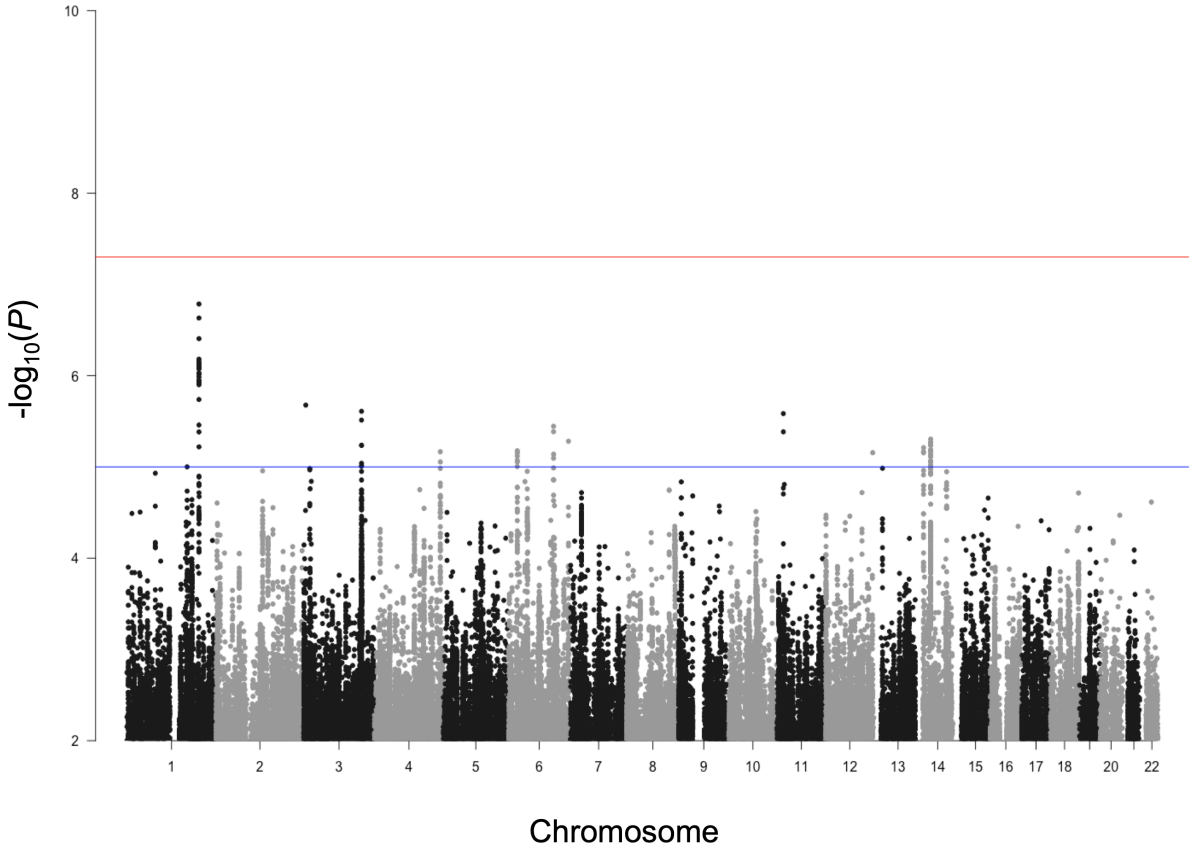

Figure 4.2: Univariable GWAS results for OS in the full patient cohort, additive model. N=1948. MAF $\geq$ 0.05, info score $>$ 0.8. MAF: Minor allele frequency. Red line: Threshold of genome-wide significance ($P < 5.0\text{x}10^{-8}$). Blue line: Threshold of suggestive association ($P < 1.0\text{x}10^{-5}$).

Table 4.1: Uni- and multivariable GWAS results for variants in the full patient cohort.

Variant	R	A	Cytoband	Univariable analyses			Multivariable analyses		
				HR	95% CI	*P*	HR	95% CI	*P*
rs706494	C	A	1q32.1	1.29	1.17-1.41	$1.6x10^{-7}$	1.26	1.13-1.41	$4.8x10^{-5}$
rs1722765	C	G	1q32.1	1.28	1.16-1.40	$2.3x10^{-7}$	1.25	1.12-1.39	$7.2x10^{-5}$
rs5020875	T	A	1q32.1	1.24	1.14-1.35	$3.9x10^{-7}$	1.19	1.08-1.31	$3.5x10^{-4}$
rs1534051	G	A	1q32.1	1.26	1.15-1.39	$6.6x10^{-7}$	1.22	1.10-1.36	$2.9x10^{-4}$
rs1779289	T	C	1q32.1	1.26	1.15-1.38	$7.0x10^{-7}$	1.21	1.09-1.35	$3.9x10^{-4}$
rs1628556	C	A	1q32.1	1.26	1.15-1.38	$7.3x10^{-7}$	1.21	1.09-1.35	$4.3x10^{-4}$
rs1722764	G	A	1q32.1	1.26	1.15-1.38	$7.6x10^{-7}$	1.22	1.09-1.36	$3.4x10^{-4}$
rs1722742	G	A	1q32.1	1.26	1.15-1.38	$7.9x10^{-7}$	1.21	1.09-1.35	$4.4x10^{-4}$
rs1779293	T	A	1q32.1	1.26	1.15-1.38	$7.9x10^{-7}$	1.21	1.09-1.35	$4.3x10^{-4}$
rs1772832	C	T	1q32.1	1.26	1.15-1.38	$8.0x10^{-7}$	1.21	1.09-1.35	$3.9x10^{-4}$
rs832150	C	A	1q32.1	1.26	1.15-1.38	$8.1x10^{-7}$	1.22	1.09-1.36	$3.4x10^{-4}$
rs1628151	G	A	1q32.1	1.26	1.15-1.38	$8.5x10^{-7}$	1.21	1.09-1.35	$4.4x10^{-4}$
rs832152	C	A	1q32.1	1.26	1.15-1.38	$9.3x10^{-7}$	1.22	1.09-1.36	$3.6x10^{-4}$
rs1722779	C	T	1q32.1	1.26	1.15-1.38	$9.6x10^{-7}$	1.21	1.09-1.35	$4.9x10^{-4}$
rs1794867	C	T	1q32.1	1.26	1.15-1.38	$9.6x10^{-7}$	1.21	1.09-1.35	$4.9x10^{-4}$
rs832148	A	G	1q32.1	1.27	1.15-1.39	$1.0x10^{-6}$	1.22	1.10-1.37	$3.7x10^{-4}$
rs1626370	G	A	1q32.1	1.26	1.15-1.38	$1.1x10^{-6}$	1.21	1.09-1.35	$5.6x10^{-4}$
rs832151	C	T	1q32.1	1.26	1.15-1.38	$1.2x10^{-6}$	1.21	1.09-1.35	$5.1x10^{-4}$
rs700472	G	T	1q32.1	1.27	1.15-1.39	$1.2x10^{-6}$	1.23	1.10-1.38	$2.6x10^{-4}$
rs700473	T	C	1q32.1	1.26	1.15-1.39	$1.2x10^{-6}$	1.24	1.11-1.38	$1.6x10^{-4}$
rs832145	T	C	1q32.1	1.27	1.15-1.39	$1.3x10^{-6}$	1.23	1.10-1.37	$3.3x10^{-4}$
rs832146	G	C	1q32.1	1.26	1.15-1.39	$1.8x10^{-6}$	1.23	1.10-1.37	$3.7x10^{-4}$
rs1400673	T	G	3p26.1	1.34	1.19-1.51	$2.1x10^{-6}$	1.19	1.03-1.37	0.02
rs1037888	A	T	3q26.1	1.19	1.11-1.28	$2.5x10^{-6}$	1.16	1.06-1.26	$6.9x10^{-4}$
rs2355023	T	C	11p15.1	0.80	0.73-0.88	$2.6x10^{-6}$	0.82	0.73-0.91	$2.0x10^{-4}$
rs9290065	C	T	3q26.1	1.19	1.11-1.28	$3.1x10^{-6}$	1.16	1.06-1.26	$6.4x10^{-4}$
rs832147	C	G	1q32.1	1.25	1.14-1.38	$3.5x10^{-6}$	1.22	1.09-1.37	$4.3x10^{-4}$
rs4896787	A	G	6q22.31	0.73	0.63-0.83	$3.6x10^{-6}$	0.74	0.63-0.86	$9.1x10^{-5}$
rs3799775	G	T	6q22.31	0.73	0.64-0.83	$4.1x10^{-6}$	0.74	0.64-0.86	$9.9x10^{-5}$
rs11024323	T	C	11p15.1	0.82	0.75-0.89	$4.1x10^{-6}$	0.83	0.75-0.91	$1.9x10^{-4}$
rs1419276	A	T	1q32.1	1.24	1.13-1.36	$4.1x10^{-6}$	1.19	1.07-1.33	$9.7x10^{-4}$
rs12886160	C	T	14q21.1	0.81	0.74-0.89	$5.0x10^{-6}$	0.84	0.76-0.93	$8.0x10^{-4}$
rs9356458	G	A	6q27	0.84	0.77-0.90	$5.2x10^{-6}$	0.76	0.70-0.83	$2.4x10^{-9}$
rs12434791	C	A	14q21.1	0.81	0.74-0.89	$5.4x10^{-6}$	0.84	0.76-0.93	$9.8x10^{-4}$
rs11157284	A	T	14q21.1	0.81	0.74-0.89	$5.5x10^{-6}$	0.84	0.76-0.93	$9.2x10^{-4}$

rs12432940	G	T	14q21.1	0.81	0.74-0.89	$5.5x10^{-6}$	0.84	0.76-0.93	$9.2x10^{-4}$
rs898681	A	G	3q26.1	1.18	1.10-1.27	$5.8x10^{-6}$	1.15	1.06-1.25	$7.0x10^{-4}$
rs898680	G	A	3q26.1	1.18	1.10-1.27	$5.8x10^{-6}$	1.16	1.06-1.26	$6.7x10^{-4}$
rs7157750	T	C	14q21.1	0.81	0.74-0.89	$5.8x10^{-6}$	0.84	0.76-0.93	$8.0x10^{-4}$
rs10800776	C	T	1q32.1	1.23	1.13-1.35	$6.0x10^{-6}$	1.19	1.07-1.32	$1.3x10^{-3}$
rs10137075	G	A	14q11.2	1.40	1.21-1.62	$6.2x10^{-6}$	1.39	1.18-1.63	$8.5x10^{-5}$
rs68089103	G	T	14q21.1	0.81	0.74-0.89	$6.5x10^{-6}$	0.84	0.76-0.94	$1.3x10^{-3}$
rs9467333	C	A	6p22.3	1.25	1.13-1.37	$6.7x10^{-6}$	1.17	1.05-1.31	$5.3x10^{-3}$
rs4437459	G	T	6p22.3	1.26	1.14-1.39	$6.8x10^{-6}$	1.24	1.10-1.39	$3.3x10^{-4}$
rs11620983	T	C	14q21.1	0.81	0.74-0.89	$6.8x10^{-6}$	0.84	0.76-0.93	$9.7x10^{-3}$
rs28407819	G	A	14q11.2	1.39	1.21-1.61	$6.8x10^{-6}$	1.39	1.18-1.63	$6.7x10^{-5}$
rs73011740	A	G	4q34.3	1.37	1.19-1.57	$6.8x10^{-6}$	1.41	1.20-1.65	$2.5x10^{-5}$
rs149908600*	T	TA	12q24.33	1.22	1.12-1.33	$7.0x10^{-6}$	1.26	1.13-1.39	$1.3x10^{-5}$
rs5742772	G	A	14q11.2	1.39	1.21-1.61	$7.0x10^{-6}$	1.39	1.18-1.63	$6.9x10^{-5}$
rs4236033	C	T	6p22.3	1.26	1.14-1.39	$7.1x10^{-6}$	1.24	1.10-1.40	$3.3x10^{-4}$
rs9767619	T	C	6q22.31	0.73	0.64-0.84	$7.3x10^{-6}$	0.73	0.63-0.86	$1.0x10^{-4}$
rs34752431*	A	AT	14q21.1	0.81	0.74-0.89	$7.3x10^{-6}$	0.81	0.73-0.90	$1.3x10^{-4}$
rs4537125	G	A	6p22.3	1.26	1.14-1.39	$7.6x10^{-6}$	1.24	1.10-1.39	$3.5x10^{-4}$
rs12436094	T	C	14q21.1	0.81	0.74-0.89	$7.7x10^{-6}$	0.85	0.76-0.94	$1.6x10^{-3}$
rs3778462	T	G	6q22.31	0.73	0.64-0.84	$8.0x10^{-6}$	0.75	0.64-0.87	$1.7x10^{-4}$
rs4236034	G	T	6p22.3	1.26	1.14-1.39	$8.5x10^{-6}$	1.23	1.10-1.39	$4.4x10^{-4}$
rs5015993	A	C	14q21.1	0.82	0.75-0.89	$8.7x10^{-6}$	0.85	0.76-0.94	$1.3x10^{-3}$
rs7144172	T	C	14q21.1	0.82	0.75-0.89	$8.7x10^{-6}$	0.84	0.76-0.94	$1.3x10^{-3}$
rs716162	T	C	14q21.1	0.82	0.75-0.89	$8.7x10^{-6}$	0.84	0.84-0.76	$1.3x10^{-3}$
rs6845276	C	T	4q34.3	1.31	1.16-1.47	$8.8x10^{-6}$	1.24	1.08-1.41	$2.6x10^{-3}$
rs4712863	T	A	6p22.3	1.26	1.14-1.39	$8.8x10^{-6}$	1.23	1.09-1.38	$5.8x10^{-4}$
rs2415728	A	C	14q21.1	0.82	0.75-0.89	$8.9x10^{-6}$	0.85	0.76-0.94	$1.3x10^{-3}$
rs980975	G	A	3q26.1	1.18	1.09-1.26	$9.1x10^{-6}$	1.15	1.06-1.25	$6.6x10^{-4}$
rs7157351	G	A	14q21.1	0.82	0.75-0.89	$9.6x10^{-6}$	0.84	0.76-0.94	$1.2x10^{-3}$
rs6769642	C	A	3q26.1	1.17	1.09-1.26	$9.8x10^{-6}$	1.15	1.06-1.25	$8.5x10^{-4}$
rs980976	T	G	3q26.1	1.18	1.09-1.26	$9.8x10^{-6}$	1.15	1.06-1.25	$7.0x10^{-4}$
rs2817749	G	A	6p22.3	1.26	1.13-1.39	$9.9x10^{-6}$	1.23	1.10-1.39	$4.3x10^{-4}$
rs2142745	A	G	1q24.2	1.34	1.18-1.53	$9.9x10^{-6}$	1.36	1.17-1.59	$8.7x10^{-5}$

*Variant is an indel. Maximum univariable n=1948, maximum multivariable n=1551. All MAFs ≥ 0.05. All variants suggestive of association with OS ($P < 1.0x10^{-5}$) in univariable analyses. R: Reference allele. A: Alternate allele (allele analysed). HR: Hazard Ratio. CI: Confidence Interval. *P*: P-value.

Further interrogation of the distribution of P-values using a Q-Q plot (Figure 4.3) found that the majority of P-values lay very close to the y=x line, indicating the absence of systematic bias. The extreme observed values that diverged from the y=x line are SNPs that are suggestive of association (Reed et al., 2015). The inflation factor estimate (λ) was λ=1.0, therefore no underlying population substructure was present.

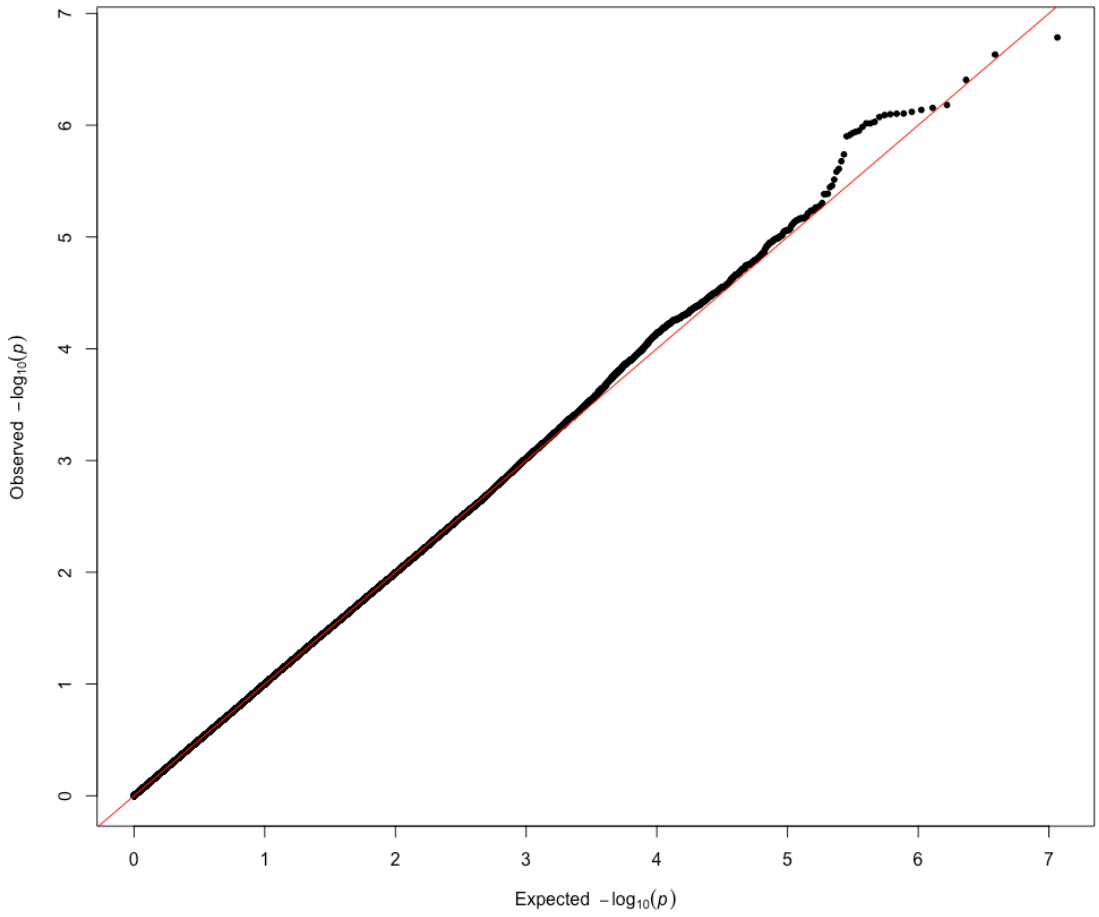

Figure 4.3: Observed versus expected P-values for univariable GWAS. MAF $\geq$ 0.05, info score $>$ 0.8. MAF: Minor allele frequency.

4.3.1.2.2 Analysis of variants suggestive of association under a multivariable model

The SNPs found to be suggestive of association through univariable GWAS analyses were interrogated further with the addition of known prognostic factors into the model. These covariates were the same as those added in multivariable analyses of somatic mutations (Table 3.7). rs9356458 at 6q27 became significant above the threshold of genome-wide significance ($P < 5.0\text{x}10^{-8}$) after the inclusion of known prognostic factors (HR 0.76, 95% CI 0.70-0.83, P=2.4x10^{-9}; Table 4.1). Given that rs9356458 only became genome-wide significant after the

inclusion of prognostic covariates, a multivariable GWAS was performed subsequently in order to determine whether any other SNPs might increase in significance to this level.

4.3.1.2.3 Multivariable GWAS analyses

multivariable GWAS analyses identified a further SNP of genome-wide significance, rs17560791, at 2q35 (HR 0.77, 95% CI 0.70-0.84, P=3.7x10^{-8} (Figure 4.4). A haplotype block of fifteen SNPs of suggestive association was also identified, with rs241477 the most significant (HR 1.40, 95% CI 1.23-1.59, P=1.8x10^{-7}). In total, SNPs at three distinct loci (6q27, 2q35 and 14q31) were identified through multivariable GWAS analyses (Table 4.1). LD analyses identified that the four other SNPs of genome-wide significance at 6q27 (rs35925426, rs11448205, rs9347113 and rs6930706) were in LD with rs9356458. One of these variants was in high LD ($r^2 > 0.8$) with rs9356458 (rs6930706, r^2=0.96), while the other three were in moderate LD (rs6930706, r^2=0.77; rs11448205, r^2=0.77 and rs9347113, r^2=0.69).

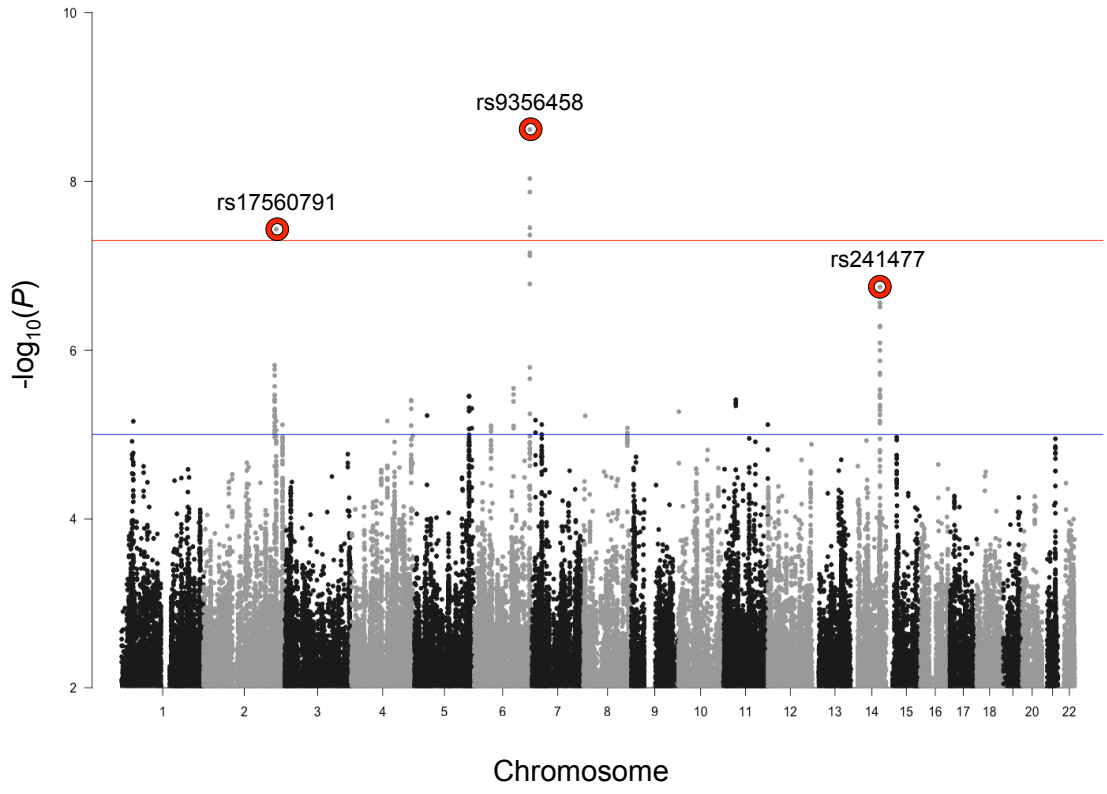

Figure 4.4: Multivariable GWAS results for OS in the full patient cohort, additive model. N=1948. MAF $\geq$ 0.05, info score $>$ 0.8. MAF: Minor allele frequency. Red line: Threshold of genome-wide significance ($P < 5.0$x10^{-8}). Blue line: Threshold of suggestive association ($P < 1.0$x10^{-5}).

4.3.1.3 Further interrogation of significant SNPs

4.3.1.3.1 Analysis under different survival measures and genetic models

In addition to visualising GWAS outputs, association analysis using other genetic models can be a useful sensitivity analysis (Reed et al., 2015). The only SNP found to be associated with OS at the level of genome-wide significance ($P < 5.0 \times 10^{-8}$) under a dominant genetic model was rs9356458 (Table 4.2). rs9356458 was also found to be suggestive of association ($P < 1.0 \times 10^{-5}$) for the diagnosis to death (DTD) survival measure under an additive and dominant genetic model for the full patient cohort (Table 4.2).

Table 4.2: Results for significantly associated variants under different analysis models.

				Overall survival			Diagnosis to death			Heterogeneity tests		
Variant	**Ref**	**Alt**	**Cytoband**	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***	**I^2 %**	**Q**	***P^{HET}***
Additive model												
rs9356458	G	A	6q27	0.76	0.70-0.83	$2.4x10^{-9}$	0.81	0.74-0.89	$4.3x10^{-6}$	0.00	0.86	0.35
rs17560791	G	C	2q35	0.77	0.70-0.84	$3.7x10^{-8}$	0.82	0.75-0.90	$4.6x10^{-5}$	2.10	1.02	0.31
rs241477	T	A	14q31.3	0.71	0.63-0.81	$1.8x10^{-7}$	0.76	0.67-0.86	$1.8x10^{-5}$	0.00	0.51	0.47
Dominant model												
rs9356458	G	A	6q27	0.69	0.61-0.78	$3.9x10^{-9}$	0.73	0.65-0.83	$5.7x10^{-7}$	0.00	0.41	0.52
rs17560791	G	C	2q35	0.68	0.59-0.78	$1.5x10^{-7}$	0.74	0.64-0.86	$6.0x10^{-5}$	0.00	0.70	0.40
rs241477	T	A	14q31.3	0.71	0.62-0.81	$9.5x10^{-7}$	0.73	0.64-0.84	$1.1x10^{-5}$	0.00	0.11	0.74

Genome-wide significant P-values are highlighted in orange. Maximum n=1551. Ref: Reference allele. Alt: Alternate allele (allele analysed). HR: Hazard Ratio. CI: Confidence Interval. *P*: P-value. OS: Overall survival. DTD: Diagnosis to death.

4.3.2 All wild type subgroup

In order to unmask the underlying prognostic effects of patients' somatic tumour profiles (identified in Chapter 3), secondary GWAS analyses were performed on a subset of the full cohort, including only those patients with *RAS* and *BRAF* wild type and MSS tumours (n=749).

4.3.2.1 Power calculations

Based on a sample size of 749 patients (499 events), GWAS analyses had $\geq$80% power to detect associations with OS for variants with HRs $\geq$ 7.41 and MAFs $\geq$ 0.01 and for variants with HRs $\geq$ 2.50 and MAFs $\geq$ 0.05 (Figure 4.5).

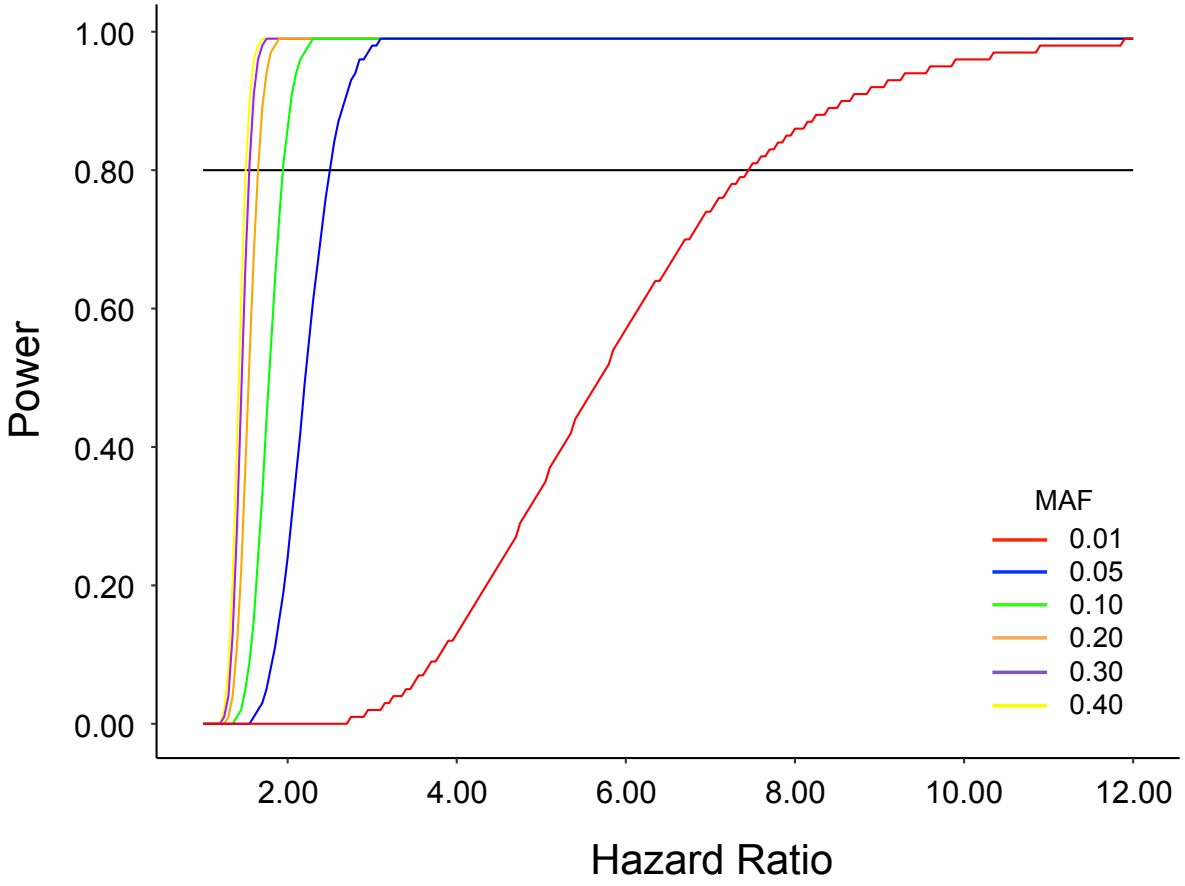

Figure 4.5: Statistical power to detect associations with OS in the all wild type subgroup. N=749 (499 events). α level: $P < 5.0\text{x}10^{-8}$. Horizontal line represents the threshold for 80% power. MAF: Minor allele frequency.

4.3.2.2 GWAS analyses

4.3.2.2.1 Univariable GWAS analyses

Under an additive genetic model for OS, 37 commonly inherited variants (MAF $\geq$ 0.05, info score $>$ 0.8) were found to have a suggestive association with survival ($P < 1.0\text{x}10^{-5}$; Figure 4.6, Table 4.3). No variants were found to be of genome-wide significance ($P < 5.0\text{x}10^{-8}$), although a peak of nine variants at 7p11.2 were extremely close to this threshold; the most significant of these being rs60455898 (P=7.8x10^{-8}; Figure 4.6, Table 4.3). LD analyses confirmed these variants as being in a haplotype block. The three SNPs previously identified through multivariable analyses of the full patient cohort did not achieve the level of suggestive association with OS in univariable analyses of the all wild type patient cohort (Figure 4.6, highlighted SNPs).

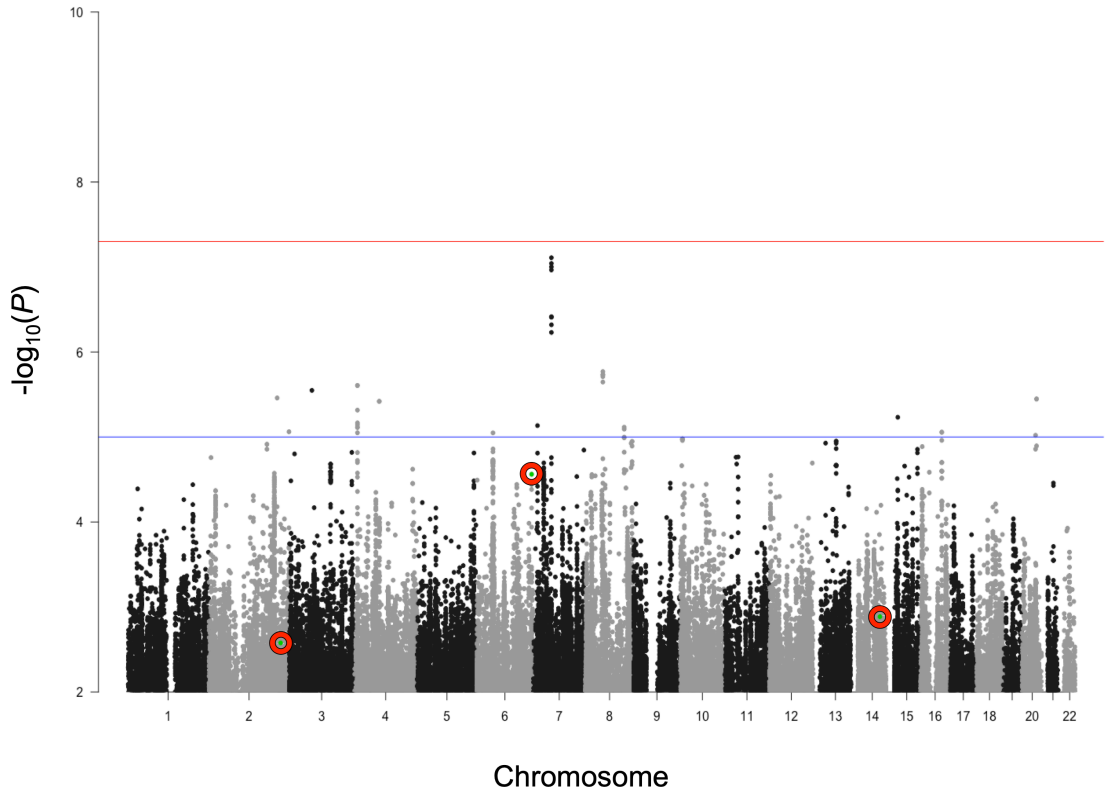

Figure 4.6: Univariable GWAS results for OS in the all wild type subgroup, additive model. N=749. MAF $\geq$ 0.05, info score $>$ 0.8. Highlighted SNPs (in red): genome-wide significant loci from full patient cohort multivariable analyses (rs17560791 at 2q35, rs9356458 at 6q27 and rs241477 at 14.31.3). MAF: Minor allele frequency. Red line: Threshold of genome-wide significance ($P < 5.0\text{x}10^{-8}$). Blue line: Threshold of suggestive association ($P < 1.0\text{x}10^{-5}$).

Table 4.3: Uni- and multivariable GWAS results for variants in the all wild type subgroup.

				Univariable analyses			Multivariable analyses		
Variant	**R**	**A**	**Cytoband**	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
rs60455898	T	C	7p11.2	1.91	1.51-2.42	$7.8x10^{-8}$	1.85	1.44-2.38	$1.5x10^{-6}$
rs6952394	T	G	7p11.2	1.91	1.50-2.41	$9.1x10^{-8}$	1.85	1.44-2.38	$1.7x10^{-6}$
rs4948063	T	C	7p11.2	1.90	1.50-2.41	$1.0x10^{-7}$	1.84	1.43-2.37	$1.8x10^{-6}$
rs4074683	C	T	7p11.2	1.94	1.52-2.48	$1.1x10^{-7}$	1.95	1.50-2.52	$4.5x10^{-7}$
rs4948064	T	C	7p11.2	1.83	1.45-2.31	$3.8x10^{-7}$	1.88	1.47-2.40	$5.9x10^{-7}$
rs55944864	A	G	7p11.2	1.83	1.45-2.31	$3.9x10^{-7}$	1.87	1.46-2.40	$6.2x10^{-7}$
rs58814124	C	T	7p11.2	1.83	1.45-2.31	$3.9x10^{-7}$	1.87	1.46-2.40	$6.2x10^{-7}$
rs35767833	T	C	7p11.2	1.82	1.44-2.30	$4.8x10^{-7}$	1.86	1.45-2.38	$8.6x10^{-7}$
rs139111483	T	C	7p11.2	1.82	1.44-2.30	$5.9x10^{-7}$	1.85	1.44-2.37	$1.1x10^{-6}$
rs200903757*	AG	A	8q11.23	1.91	1.46-2.48	$1.7x10^{-6}$	1.85	1.39-2.44	$2.0x10^{-5}$
rs10504132	C	T	8q11.23	1.90	1.46-2.48	$1.8x10^{-6}$	1.84	1.39-2.43	$2.3x10^{-5}$
rs7012946	G	A	8q11.23	1.90	1.46-2.48	$1.8x10^{-6}$	1.84	1.39-2.43	$2.3x10^{-5}$
rs10104368	C	A	8q11.23	1.91	1.46-2.49	$2.0x10^{-6}$	1.84	1.29-2.43	$2.2x10^{-5}$
rs13281179	G	C	8q11.23	1.90	1.46-2.48	$2.3x10^{-6}$	1.83	1.38-2.42	$2.5x10^{-5}$
rs56880996	G	C	4p16.1	1.46	1.25-1.71	$2.5x10^{-6}$	1.36	1.15-1.62	$3.8x10^{-4}$
rs56977009	A	T	4p16.1	1.46	1.25-1.71	$2.5x10^{-6}$	1.36	1.15-1.62	$3.8x10^{-4}$
rs112083640	C	A	3p14.1	1.75	1.39-2.21	$2.8x10^{-6}$	1.72	1.34-2.21	$1.8x10^{-5}$
rs200204853*	DEL 14 BP	A	2q33.3	1.66	1.34-2.05	$3.5x10^{-6}$	1.64	1.31-2.05	$1.3x10^{-5}$
rs11908352	C	A	20q13.12	1.51	1.27-1.79	$3.6x10^{-6}$	1.30	1.08-1.57	$5.5x10^{-3}$
rs6124764	T	A	20q13.12	1.51	1.27-1.79	$3.6x10^{-6}$	1.30	1.08-1.57	$5.5x10^{-3}$
rs114841535	T	C	4q13.3	0.53	0.40-0.69	$3.8x10^{-6}$	0.54	0.41-0.72	$2.5x10^{-5}$
rs116033880	G	A	4q13.3	0.53	0.40-0.69	$3.8x10^{-6}$	0.54	0.41-0.72	$2.5x10^{-5}$
rs34230659*	TC	T	4p16.1	1.44	1.23-1.69	$4.8x10^{-6}$	1.37	1.15-1.62	$3.0x10^{-4}$
rs7179722	T	A	15q14	0.73	0.64-0.84	$5.9x10^{-6}$	0.75	0.65-0.87	$9.5x10^{-5}$
rs1878943	C	T	4p16.1	1.42	1.22-1.66	$6.8x10^{-6}$	1.36	1.15-1.60	$3.4x10^{-4}$
rs61166583	G	A	4p16.1	1.43	1.22-1.67	$7.2x10^{-6}$	1.36	1.15-1.61	$4.2x10^{-4}$
rs2528512	G	A	7p21.2	1.44	1.23-1.68	$7.3x10^{-6}$	1.49	1.26-1.75	$2.7x10^{-6}$
rs2317214	A	G	8q23.3	1.48	1.24-1.75	$7.7x10^{-6}$	1.40	1.18-1.68	$1.8x10^{-4}$
rs1818669	C	G	4p16.1	1.42	1.22-1.65	$7.7x10^{-6}$	1.35	1.14-1.60	$3.9x10^{-4}$
rs73090773	C	A	4p16.1	1.42	1.22-1.65	$7.7x10^{-6}$	1.35	1.14-1.60	$3.9x10^{-4}$
rs5894337*	G	GT	8q23.3	1.50	1.25-1.79	$8.1x10^{-6}$	1.42	1.18-1.71	$2.3x10^{-4}$
rs35720921*	T	TA	2q37.3	1.46	1.23-1.72	$8.7x10^{-6}$	1.58	1.32-1.89	$6.1x10^{-7}$
rs10500523	T	G	16q21	1.61	1.31-1.99	$8.8x10^{-6}$	1.60	1.29-1.98	$1.7x10^{-5}$

rs17526038	A	C	16q21	1.61	1.31-1.99	$8.8x10^{-6}$	1.60	1.29-1.98	$1.7x10^{-5}$
rs73090775	C	T	4p16.1	1.42	1.22-1.65	$8.9x10^{-6}$	1.35	1.14-1.60	$4.3x10^{-4}$
rs2448702	G	A	6p12.3	1.32	1.17-1.50	$8.9x10^{-6}$	1.31	1.15-1.49	$3.8x10^{-5}$
rs881497	C	T	20q12	1.41	1.21-1.64	$9.5x10^{-6}$	1.47	1.25-1.72	$2.9x10^{-6}$

*Variant is an indel. Maximum univariable n=749, maximum multivariable n=740. All MAFs ≥ 0.05. All variants suggestive of association with OS ($P < 1.0x10^{-5}$) in univariable analyses. R: Reference allele. A: Alternate allele (allele analysed). HR: Hazard Ratio. CI: Confidence interval. *P*: P-value. DEL: Deletion. BP: Base-pairs.

4.3.2.2.2 Analysis of variants suggestive of association under a multivariable model

The SNPs found to be suggestive of association through univariable GWAS analyses were interrogated further with the addition of known prognostic factors into the model. These covariates were the same as those added in multivariable analyses of somatic mutations (Table 3.7). None of these SNPs achieved genome-wide significance under a multivariable model, although a full multivariable GWAS was then performed in order to ascertain whether any of the genome-wide significant SNPs identified through multivariable GWAS analyses of the full patient cohort achieved a level of suggestive association in the all wild type subgroup.

4.3.2.2.3 Multivariable GWAS analyses

There was support for rs9356458 (6q27) at the level of suggestive association ($P < 1.0\text{x}10^{-5}$), but not for rs17560791 (2q35) or rs241477 (14q31.3; Figure 4.7). The peak of variants approaching genome-wide significance remained suggestive of association in multivariable analyses, of which rs4074683 was the most significant SNP.

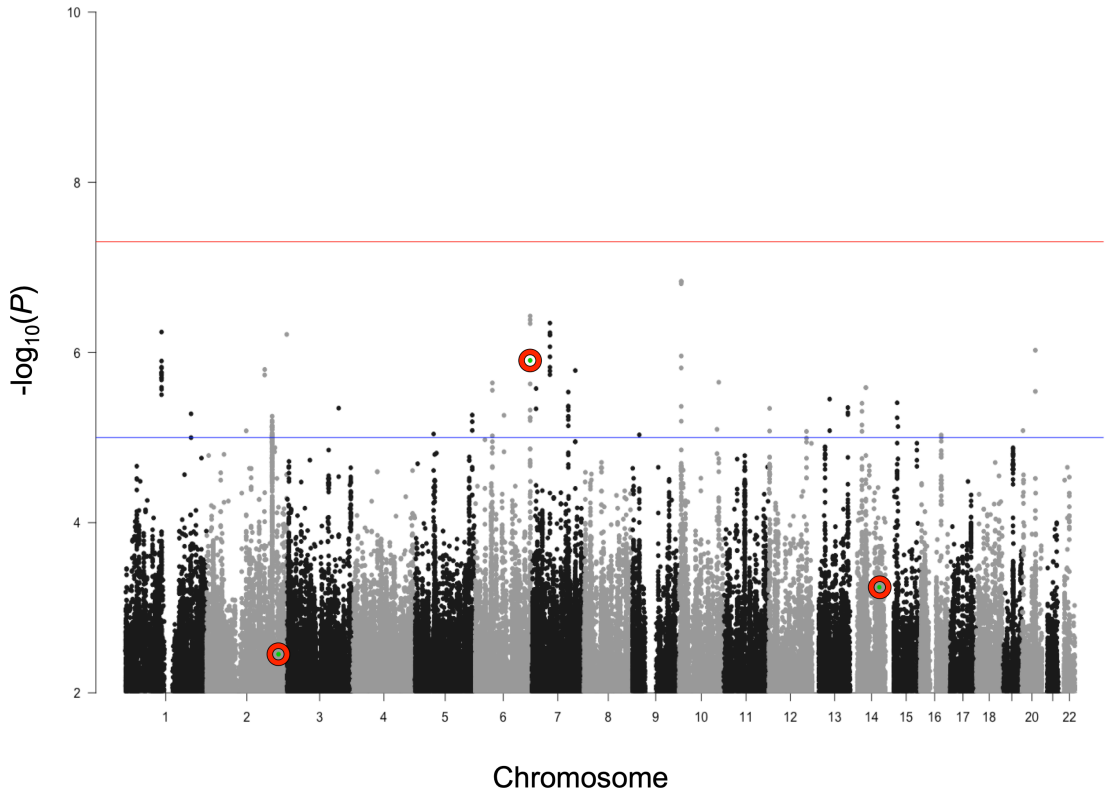

Figure 4.7: Multivariable GWAS results for OS in the all wild type subgroup, additive model. N=749. MAF ≥ 0.05, info score> 0.8. Highlighted SNPs (in red): genome-wide significant loci from full patient cohort multivariable analyses (rs17560791 at 2q35, rs9356458 at 6q27 and rs241477 at 14.31.3). MAF: Minor allele frequency. Red line: Threshold of genome-wide significance ($P < 5.0\text{x}10^{-8}$). Blue line: Threshold of suggestive association ($P < 1.0\text{x}10^{-5}$).

4.3.3 Support for previously proposed and established prognostic biomarkers

The two SNPs previously found to be robustly associated with CRC survival, rs9929218 and rs209489 (Smith et al., 2015; Phipps et al., 2016), were not found to be significantly associated with OS under an additive genetic model (rs9929218: HR 1.11, 95% CI 1.02-1.20, P=1.2x10^{-2}; rs209489: HR 1.01, 95% CI 0.88-1.16, P=0.89). However, under the recessive model, rs9929218 was found to be suggestive of association with OS (HR 1.42, 95% CI 1.19-1.69, P=9.6x10^{-5}; Table 4.4). No support was found for any other SNPs previously reported to be associated with survival in the literature (Passarelli et al., 2011; Dai et al., 2012; Phipps et al., 2012; Garcia-Albeniz et al., 2013; Abuli et al., 2013; Takatsuno et al., 2013; Morris et al., 2015) to a level of significance suggestive of association, although support existed for rs10795668 at a level of nominal significance ($P < 0.05$; Table 4.4).

Table 4.4: Results for previously proposed CRC prognostic loci in COIN & COIN-B.

					OS (previous study)			OS (COIN & COIN-B)		
Variant	**Cytoband**	**Study**	**Ref**	**Alt**	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
rs209489*	6p12.1	Phipps et al. 2016	A	C	1.80	1.50 - 2.10	$3.7x10^{-9}$	0.96	0.82 - 1.12	0.57
rs6983267	8q24.21	Dai et al. 2012	G	T	0.24	0.06 - 0.88	0.03	1.00	0.88 - 1.14	0.99
rs10795668	10p14	Phipps et al. 2012	G	A	1.14	1.02 - 1.28	0.02	0.91	0.83 - 0.99	0.04
rs4444235	14q22.2	Morris et al. 2015	T	C	1.13	1.05 - 1.22	$1.0x10^{-3}$	1.00	0.92 - 1.08	0.96
rs9929218*	16q22.1	Smith et al. 2015	G	A	1.28	1.14 - 1.43	$2.2x10^{-5}$	1.42	1.19 - 1.69	$9.6x10^{-5}$
rs4939827	18q21.1	Phipps et al. 2012	T	C	1.13	1.01 - 1.25	0.03	0.98	0.91 - 1.07	0.69
rs961253	20p12.3	Dai et al. 2012	C	A	0.24	0.09 - 0.68	$7.0x10^{-3}$	0.99	0.88 - 1.12	0.93
rs4925386	20q13.33	Phipps et al. 2012	C	T	0.88	0.78 - 0.98	0.03	1.13	0.91 - 1.39	0.27

*Confirmed as a prognostic marker for CRC using an independent validation cohort (data shown from combined discovery and follow-up cohorts). rs209489 and rs9929218 were identified as a prognostic markers through a GWAS and candidate gene study, respectively. The COIN and COIN-B cohorts were used in the training and validation phases, respectively in the discovery of rs9929218, and the validation phase in the discovery of rs209489. rs209489, rs10795668, rs4444235, rs4939827 and rs4925386 were analysed using an additive model, rs6983267 and rs961253 using a dominant model and rs9929218 using a recessive model. COIN and COIN-B data was taken from multivariable analyses. Results from previous studies have been adjusted where appropriate to ensure the same allele was analysed. Ref: Reference allele. Alt: Alternative allele (allele analysed). HR: Hazard ratio. CI: Confidence interval. *P*: P-value.

4.4 Discussion

4.4.1 The identification of novel germline prognostic biomarkers for CRC

The search for novel germline prognostic biomarkers for CRC has to date yielded very few robustly associated variants. This is likely to be due in part to the small sample sizes used in these studies and the associated lack of sufficient statistical power required to detect variants with true prognostic effects (Smith et al., 2015; Phipps et al., 2016). Through analysis of germline data from the largest clinical trial of mCRC patients to date (the combined COIN and COIN-B cohorts), the work in this chapter has aimed to find novel germline prognostic biomarkers that previous studies were unable to detect.

Initially, univariable GWAS analyses were performed using strict quality control thresholds for MAF (>0.05) and info score (>0.8) in order to ensure that the results obtained were as robust as possible. No variants were identified at the level of genome-wide significance ($P<5.0\text{x}10^{-8}$) through univariable GWAS analyses, although 68 variants were found to be above the level of suggestive association ($P<1.0\text{x}10^{-5}$). These variants were analysed further using a multivariable model in order to ascertain whether the inclusion of known clinical and prognostic factors would impact their levels of significance. Following the inclusion of these factors into the model, one SNP (rs9356458 at 6q27) increased in significance from $P=5.2\text{x}10^{-6}$ to $P=2.4\text{x}10^{-9}$. The covariates that had the greatest impact on this increase in significance were WHO PS ($P=7.0\text{x}10^{-5}$), WBC count ($P=4.9\text{x}10^{-5}$), ALKP levels ($P=1.2\text{x}10^{-6}$), PLT count ($P=2.3\text{x}10^{-7}$), number of metastatic sites ($P=1.2\text{x}10^{-4}$), *KRAS* ($P=1.8\text{x}10^{-9}$), *BRAF* ($P<2.0\text{x}10^{-16}$) and *NRAS* status ($P=4.5\text{x}10^{-3}$) and the rs9929218 genotype ($P=6.8\text{x}10^{-6}$).

The identification of this genome-wide significant SNP led to the decision to perform a full multivariable GWAS, which identified another SNP at the level of genome-wide signficance (rs17560791 at 2q35) and a haplotype block of fifteen SNPs at 14q31.3, of which the most significant SNP was rs241477. While the possibility that rs241477 represents a false-positive result is higher than for the two SNPs of genome-wide significance, the support of the other fourteen variants identified as being in LD with this SNP suggests that this observation could be a genuine association. rs9356458, rs17560791 and rs241477 had HRs <1, indicating that the alternate allele of these variants may confer a median increase in life expectancy on those patients who carry them over those that do not.

Although the introduction of covariates into the model reduced the sample size, there was still adequate statistical power to identify variants associated with OS at a level of genome-wide significance. While univariable analyses are useful for describing survival with respect to the factor under investigation, they essentially ignore the impact of any other factors that may have an effect on the outcome (Bradburn et al., 2003). It is for this reason that clinical studies often use a multivariable analysis model to assess survival with respect to several factors simultaneously (Bradburn et al., 2003).

4.4.2 Unmasking the effects of underlying somatic prognostic factors

The results of this chapter show that unmasking the effects of the underlying somatic mutational profile of patients' tumours has an effect on the germline variants identified as being associated with CRC prognosis. To the candidate's knowledge, no other GWAS has been performed that takes this into account. This is likely to be due to the majority of other studies not having the wealth of clinical molecular data that was collected as part of the COIN and COIN-B trials. For those studies that have collected some of this data, their sample sizes are likely to be too small to allow sufficient power with which to subset the datasets in this way and still be able to detect true findings.

The results of univariable GWAS of the all wild type subgroup showed that when patients with somatic mutations in the *KRAS*, *BRAF* and *NRAS* oncogenes or whose tumours were MSI-positive were excluded, no support was found for the three highly significant variants found through GWAS analyses of the full patient cohort, although this may be due to the lower power of this subset relative to the full patient cohort. There were, however, 37 variants that were found to be suggestive of association with CRC prognosis, although none of these were significant to the level of genome-wide significance. A peak of nine SNPs were identified at 7p11.2, of which the most significant SNP was rs60455898 (P=7.8x10^{-8}). When these variants were analysed under a multivariable model, none of these variants were found to increase in significance to the level of genome-wide significance, although the peak of nine SNPs at 7p11.2 remained suggestive of association, with the most significant SNP changing to rs4074683 (P=4.5x10^{-7}) under a multivariable model.

Following this, a full multivariable GWAS was performed in order to determine how other variants changed in significance, and to see what support there was for the three SNPs identified through multivariable GWAS analyses of the full patient cohort in the all wild type subgroup. There was support for rs9356458 (6q27) at the level of suggestive association, but not for rs17560791 (2q35) or rs241477 (14q31.3). These results indicate that the germline variants associated with CRC prognosis appear to differ between patients whose tumours have somatic mutations in *KRAS*, *BRAF* and *NRAS* and are MSI-positive, and those that are not. A recent study into how germline susceptibility variants impact clinical outcome and therapeutic strategies has shown that underlying germline genetic alterations mold the tumour somatic alteration landscape of patients with Stage III CRC, which supports the results shown here (Lin et al., 2019).

4.4.3 Support for previously established and proposed germline prognostic biomarkers for CRC

There was no support for the majority of variants that have previously been proposed as prognostic biomarkers for CRC. This may largely be due to these variants being only nominally associated with survival to CRC in their respective studies, and also possibly due to differences in treatment and clinicopathological stage of patients included in each study. rs9929218 was identified through a candidate gene study that used a recessive model and COIN and COIN-B

data in the training and validation phases, respectively (Smith et al., 2015). Therefore, it is to be expected that this variant was shown to be significantly associated with survival in this study.

4.4.4 Conclusion

The results of this chapter have identified four SNPs associated with survival in CRC (two at a level of genome-wide significance and two at a level of suggestive association) through multivariable GWAS analyses. In order to gain an initial insight into the biological implications of these SNPs, further work is required using *in silico* techniques to identify the underlying mechanisms these SNPs may impact. This is the focus of the following chapter.

Chapter 5

In silico functional investigation of potential prognostic variants

5.1 Introduction

Although the GWAS approach has proven useful for identifying associations between variants and traits, a GWAS cannot determine the functional mechanisms underlying the observed associations (Manolio, 2010). Therefore, in order to shed more light on the potential relationship between variants and specific traits identified through a GWAS, further analyses are required (Carethers et al., 2015). The contextualisation of GWAS results is important as this often provides more meaningful insights into disease pathogenesis (Reed et al., 2015). The identification of the biological mechanisms a variant may influence has vast potential for clinical utility (Erichsen and Chanock, 2004). The functional analysis of SNPs and their underlying effects is therefore crucial in order to create a better understanding of their role in complex diseases and develop new markers for medical testing (Shastry, 2009).

Germline variants can have a range of functional consequences, including alterations to protein coding (Manolio, 2010) and influencing gene expression (Porcu et al., 2019). Variants that explain a fraction of the genetic variance of a gene expression phenotype are known as eQTLs (Nicolae et al., 2010; Nica and Dermitzakis, 2013). Based on high-throughput genotyping data, it has been demonstrated that trait-associated loci are three times more likely to be eQTLs than other SNPs (Nica et al., 2010; Nicolae et al., 2010), suggesting that many SNP-trait associations act through influencing gene expression (Porcu et al., 2019).

This chapter is concerned with identifying the putative functional consequences of three SNPs found to be associated with survival through multivariable GWAS analyses in the full patient cohort (two at the level of genome-wide significance; rs9356458 at 6q27 [P=2.4x10^{-9}] and rs17560791 at 2q35 [P=3.7x10^{-8}], and one suggestive of association; rs241477 at 14q31.3 [P=1.8x10^{-7}]), and a further SNP representing a haplotype block of SNPs that was found suggestive of association in both univariable and multivariable analyses of the all wild type subgroup (rs4074683 at 7p11.2 [P=4.5x10^{-7}]; Chapter 4), to better understand their potential role in CRC prognosis.

5.1.1 Aims and objectives

- To analyse whether any genes are in the local region of the four SNPs identified through multivariable analyses using regional association plots
- To ascertain whether any of these SNPs influence gene expression through conducting eQTL analyses
- To gain an initial insight into the underlying mechanisms these variants may affect through a literature search using the PubMed database

5.2 Materials and methods

5.2.1 Study design and statistical analysis methods

5.2.1.1 Selection of SNPs for functional interrogation

Four SNPs from the GWAS analyses were selected to be analysed using *in silico* methods. Two SNPs (rs9356458 at 6q27 and rs17560791 at 2q35) were identified as being associated with CRC prognosis at the level of genome-wide significance ($P < 5.0 \times 10^{-8}$) in multivariable analyses of the full patient cohort. One further SNP in the full patient cohort represented a haplotype block of SNPs suggestive of association ($P < 1.0 \times 10^{-5}$) in multivariable analyses of the full patient cohort (rs241477 at 14q31.3). One SNP in the all wild type subgroup (rs4074683 at 7p11.2) represented a haplotype block of SNPs that were found to be suggestive of association in both univariable and multivariable analyses.

5.2.1.2 Regional association analyses

Regional association analyses were performed to graphically represent the LD between variants at specific loci and to identify genes in the local region of variants identified through GWAS analyses. Results for the variants in the local region to SNPs of interest were exported from R to a text file for use with the online regional association mapping tool LocusZoom, which was used to create all regional association plots. The LD of variants in the region was calculated with respect to the most significant SNP in the region (the sentinel SNP). These figures were modified in Microsoft PowerPoint prior to inclusion in this thesis.

5.2.1.3 eQTL analyses

5.2.1.3.1 Patient and specimen characteristics

The eQTL analyses in this chapter have used data from 838 donors in the GTEx Project database (https://gtexportal.org/home/). 32.9% of donors were female; 84.6% were white, 12.9% African American, 1.3% Asian, 0.2% American Indian and 1.1% unknown. The age of donors ranged from 20 to 79 years. In total, eQTL data are provided for 15201 tissue samples. A detailed description of assay methods for GTEx Project tissue samples can be found at https://gtexportal.org/home/documentationPage. In brief, a combination of Next Generation Sequencing (NGS) and gene expression arrays were used in the collection of expression data. Genotype data collection was performed using a combination of SNP arrays and NGS.

For each SNP, data from the GTEx Project database was used to ascertain whether it was an eQTL for any genes in any of 49 distinct tissue types (although many of these tissue types were not directly relevant to this study, all tissues in the GTEx Project database were tested to ensure eQTL analyses remained unbiased). Multi-tissue eQTL plots were then created using the available online graphical tools. These figures were modified in Microsoft PowerPoint prior to inclusion in this thesis. The Bonferroni correction for multiple testing was employed for eQTL

analyses to correct for a maximum of 49 different tissues being tested. The significance level of an eQTL would therefore have to be $P < 1.0\text{x}10^{-3}$ in order to retain statistical significance.

5.2.1.4 Further interrogation of eQTL-associated genes

Further interrogation of eQTL-associated genes was performed using the PubMed database (https://www.ncbi.nlm.nih.gov/pubmed/), which was used to gain an insight into the underlying biological functions associated with eQTLs and the genes they influence.

5.3 Results

5.3.1 Regional association analyses

Ten genes were found within 400 kilobases (Kb) up- and down-stream of rs9356458; *Long Intergenic Non-Protein Coding RNAs 473* and *602* (*LINC00473* and *LINC00602*), *Proline Rich 18* (*PRR18*), *SFT Domain Containing 1* (*SFT2D1*), *Mitochondrial Pyruvate Carrier 1* (*MPC1*), *Ribosomal Protein S6 Kinase A2* (*RPS6KA2*), *Ribosomal Protein S6 Kinase A2 Intronic Transcript 1* (*RPSKA2-IT1*), *MicroRNA 1913* (*MIR1913*) and the uncharacterised gene *LOC100289495*. The most significant variants at this locus were found to be in LD with rs9356458 (Figure 5.1A).

Nine genes were found within 400 Kb up- and down-stream of rs17560791; *X-Ray Repair Cross Complementing 5* (*XRCC5*), *Long Intergenic Non-Protein Coding RNAs 1966* and *1280* (*LINC01963* [or *PKI55*] and *LINC01280*), *Membrane Associated Ring-CH-Type Finger 4* (*MARCH4*), *SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A Like 1* (*SMARCAL1*), *Ribosomal Protein L37A* (*RPL37A*), *Insulin Like Growth Factor Binding Proteins 2* and *5* (*IGFBP2* and *IGFBP5*) and *Transition Protein 1* (*TNP1*) (Figure 5.1B).

A single uncharacterised gene, *LOC283585*, was found within 400 Kb up- and down-stream of rs241477 (Figure 5.2A). Twelve genes were identified in the local region of rs4074683; *LanC Like 2* (*LANCL2*), *VOPP1 WW Domain Binding Protein* (*VOPP1*), *FKBP Prolyl Isomerase 9 Pseudogene 1* (*FKBP9P1*), *Septin 14*(*SEPT14*), *Zinc Finger Protein 713* (*ZNF713*), *Phosphoserine Phosphatase* (*PSPH*), *Mitochondrial Ribosomal Protein S17* (*MRPS17*), *Chaperonin Containing TCP1 Subunit 6A* (*CCT6A*), *Glioblastoma Amplified Sequence* (*GBAS*), *Small Nucleolar RNA H/ACA Box 15* (*SNORA15*), *Sulfatase Modifying Factor 2* (*SUMF2*) and *Phosphorylase Kinase Catalytic Subunit Gamma 1* (*PHKG1*) (Figure 5.2B).

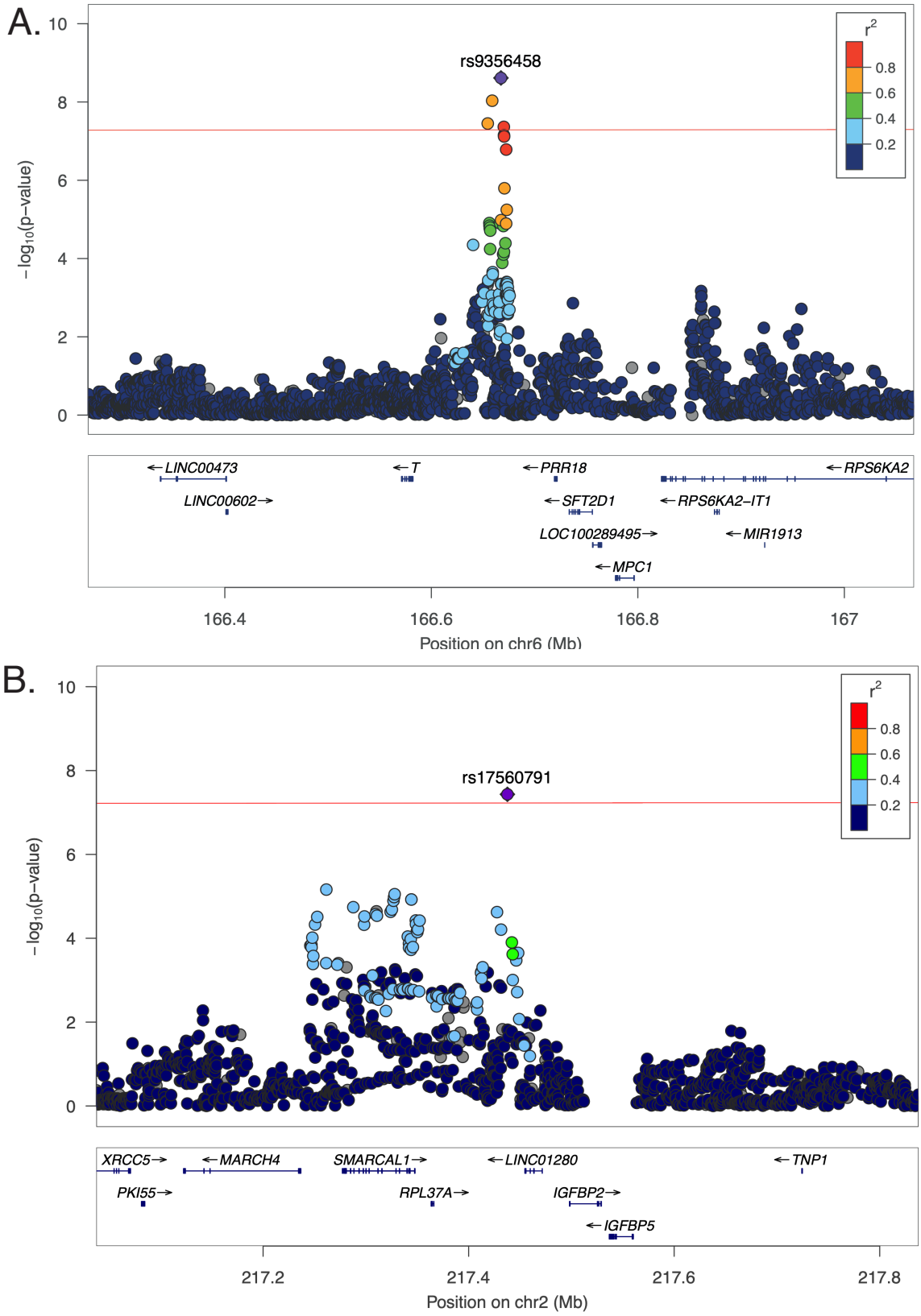

Figure 5.1: Regional associations for genome-wide significant SNPs associated with OS. A: rs9356458. B: rs17560791. Red line: genome-wide significance threshold ($P < 5.0 \times 10^{-8}$). No LD information was available for SNPs highlighted in grey. r^2: magnitude of LD. P: P-value. Mb: Megabase.

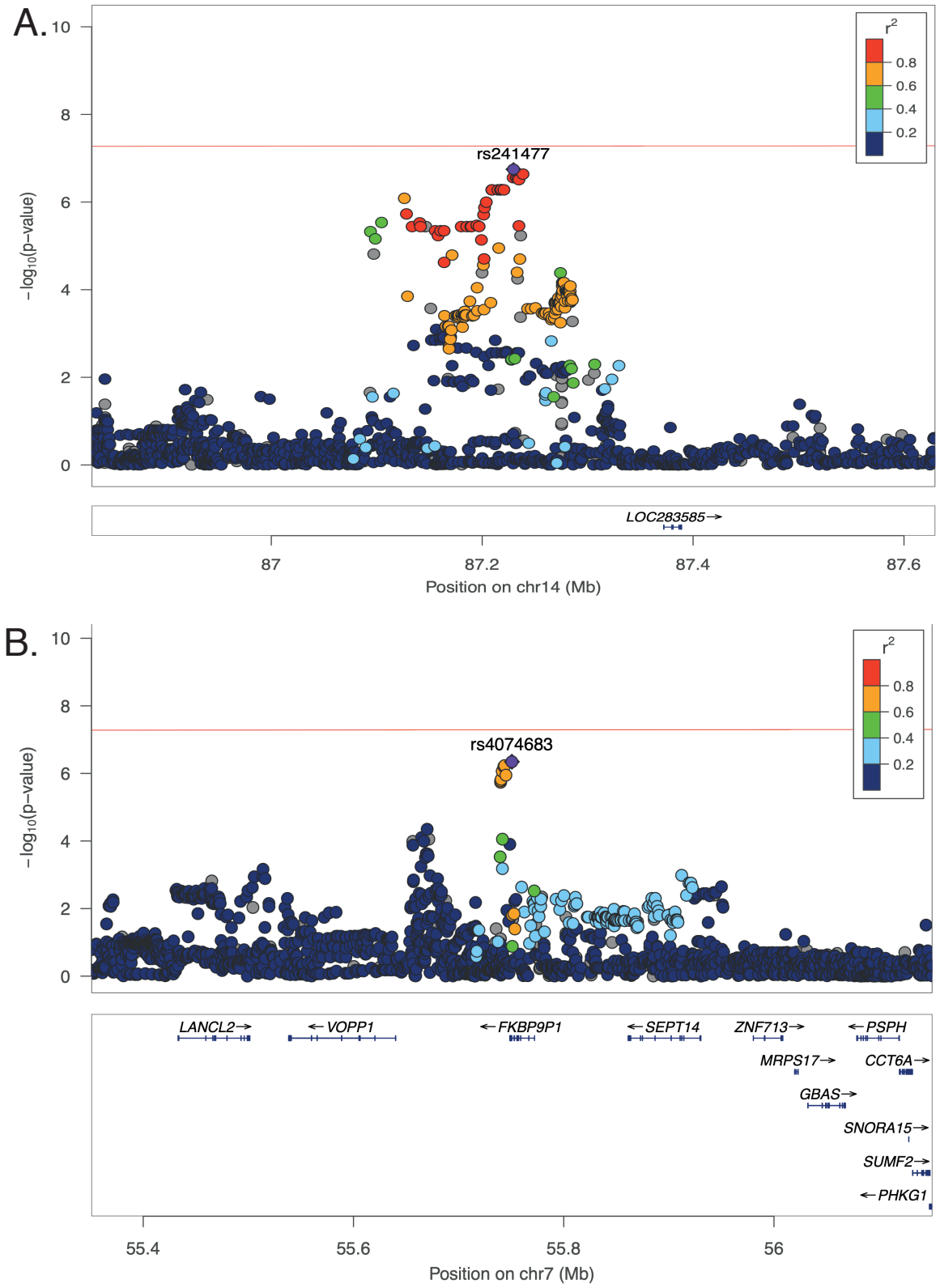

Figure 5.2: Regional associations for SNPs suggestive of association with OS. A: rs241477. B: rs4074683. Red line: genome-wide significance threshold ($P < 5.0\text{x}10^{-8}$). No LD information was available for SNPs highlighted in grey. r^2: magnitude of LD. P: P-value. Mb: Megabase.

5.3.2 eQTL analyses

A number of significant eQTL effects were identified for rs9356458, rs17560791 and rs4074683. rs9356458 was identified as an eQTL for *MPC1* in whole blood (P=9.0x10^{-8}, P=4.4x10^{-6} after the Bonferroni correction for multiple testing, Figure 5.3).

rs17506791 was found to be an eQTL for *IGFBP2* in cultured fibroblast cells (P=3.8x10^{-10}, P=1.9x10^{-8} after correction, Figure 5.4A), *SMARCAL1* in cultured fibroblast cells (P=4.2x10^{-9}, P=2.1x10^{-7} after correction, the oesophagus (P=1.1x10^{-4}, P=5.4x10^{-3} after correction), skin (both sun exposed and not sun exposed, [P=1.5x10^{-3} and P=0.02, after correction P=0.07 and P=0.98, respectively]) and heart (left ventricle, P=0.02, P=0.98 after correction, Figure 5.4B). rs17560791 was also identified as an eQTL for *AC098820.4* in 17 tissues (1.8x10$^{-5}\leq P\leq$ 0.04), including the transverse colon (P=0.01, Figure 5.4C). The tissues that remained significant following correction for multiple testing were the thyroid (P=8.8x10^{-4}), oesophagus (P=0.01), brain - nucleus accumbens (basal ganglia) (P=0.03) and spleen (P=0.03).

rs4074683 was found to be an eQTL for *PSPHP1* in 43 tissues (1.9x10$^{-14}\leq P\leq$ 0.04, 32 after correction for multiple testing; 9.3x10$^{-13}\leq P\leq$ 0.02, Figure 5.5A), *PSPH* in 34 tissues (5.6x10$^{-13}\leq P\leq$ 0.03, 21 after correction; 2.7x10$^{-11}\leq P\leq$ 0.05, Figure 5.5B), *NUPR2* in 34 tissues (2.1x10$^{-11}\leq P\leq$ 0.04, 18 after correction; 1.0x10$^{-9}\leq P\leq$ 0.03, Figure 5.5C), *RP11-613E4.4* in five tissues (3.0x10$^{-4}\leq P\leq$ 0.03, after correction only the tibial nerve was significant [P=0.01]) and *TUBBP6* and *RP11-310H4.6* in the testis (P=1.4x10^{-6} and P=3.2x10^{-6} [P=6.9x10^{-5} and P=1.6x10^{-4} after correction], respectively). In terms of colon-specific associations, rs4074683 was identified to be a significant eQTL for *PSPHP1*, *NUPR2* and *PSPH* in both the transverse (P=5.5x10^{-9}, P=5.1x10^{-6} and P=5.3x10^{-4} [P=2.7x10^{-7}, P=2.5x0^{-4} and P=0.03 after correction], respectively) and sigmoid colon (P=1.2x10^{-6}, P=8.5x10^{-5} and P=1.5x10^{-3} [P=5.9x10^{-5}, P=4.2x10^{-3} and P=0.07 after correction], respectively). rs241477 was not found to be a significant eQTL for any genes in any of the tissues analysed.

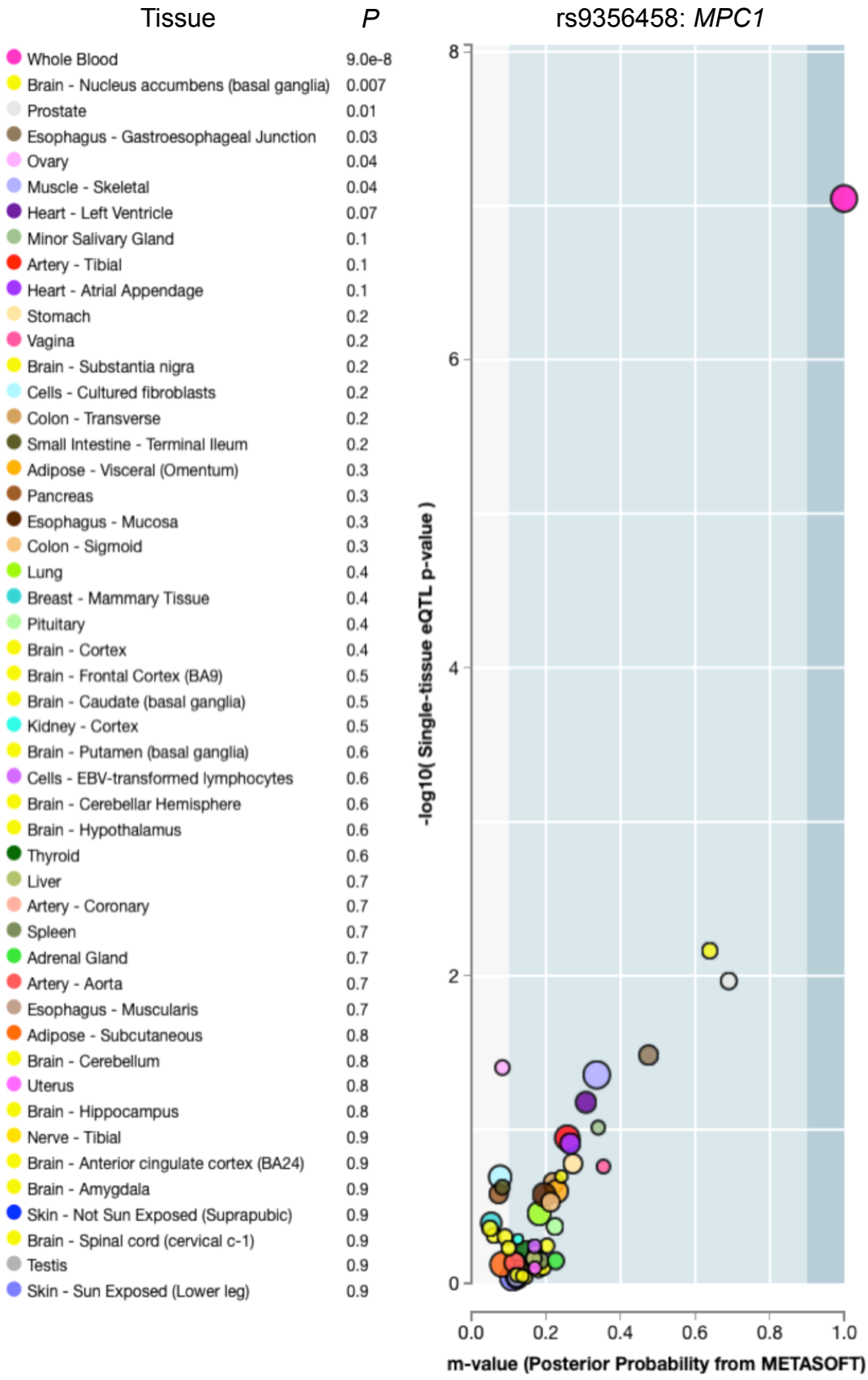

Figure 5.3: Multi-tissue eQTL associations for rs9356458. *P*: Single-tissue eQTL P-value.

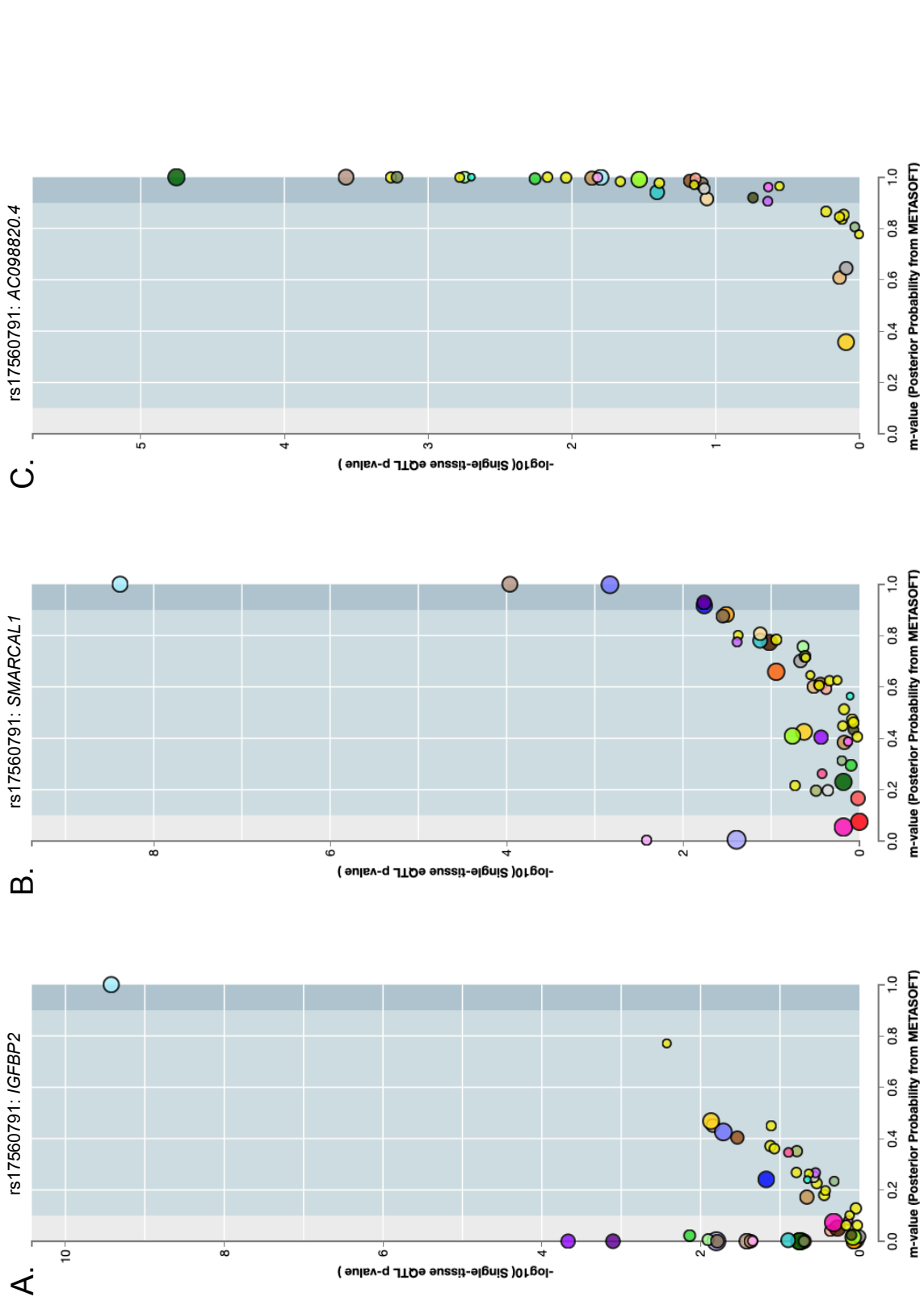

Figure 5.4: Multi-tissue eQTL associations for rs17560791. A: *IGFBP2*. B: *SMARCAL1*. C: *AC098820.4*. Tissue key as per Figure 5.3.

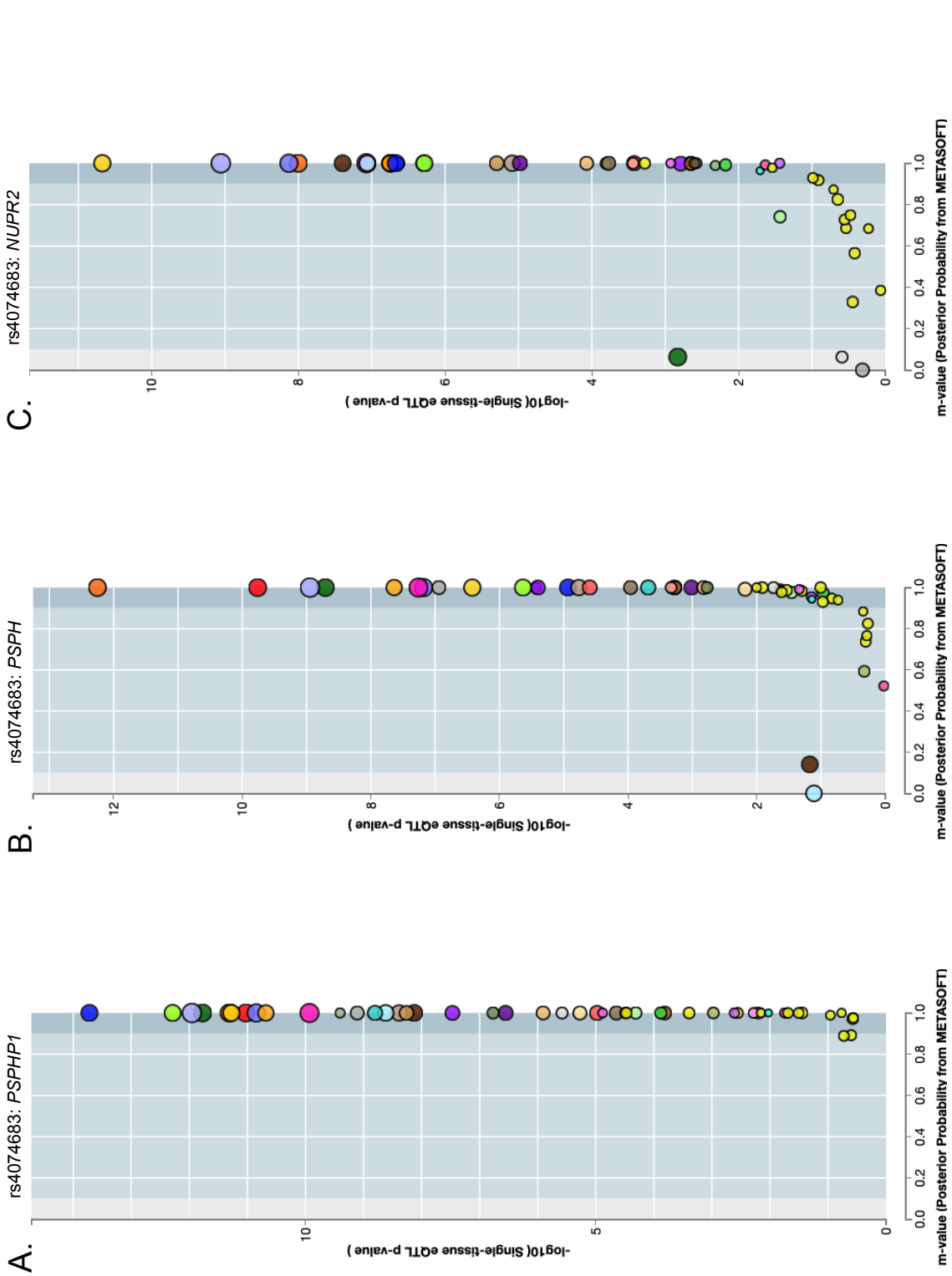

Figure 5.5: Multi-tissue eQTL associations for rs4074683. A: *PSPHP1*. B: *PSPH*. C: *NUPR2*. Tissue key as per Figure 5.3.

5.3.3 Further investigation of eQTL-associated genes

The genes found to be associated with eQTLs were investigated further through a literature search using the PubMed database (Table 5.1).

Table 5.1: Evidence for tumourigenic and survival effects of eQTL-associated genes.

Cancer	Expression	Tumourigenic effect	Cell survival effect	Reference
		MPC1		
Colorectal	↓	+	-	Schell et al. 2014
Colorectal	↓	NA	NA	Sandoval et al. 2017
Bladder	↓	+	-	Schell et al. 2014
Brain	↓	+	-	Schell et al. 2014
Kidney	↓	+	-	Tang et al. 2019
Lung	↓	+	+	Zou et al. 2019
Lung	↓	+	-	Schell et al. 2014
Prostate	↓	+	-	Wang et al. 2016
Prostate	↑	NA	+	Li et al. 2016b
Stomach	↑	-	NA	Zhou et al. 2017b
		IGFBP2		
Colorectal	↑	NA	-	el Atiq et al. 1994; Liou et al. 2010
Breast	↑	+	NA	Busund et al. 2005
Glioma	↑	+	NA	Fukushima and Kataoka 2007
Pancreatic	↑	+	-	Gao et al. 2016
Prostate	↑	+	NA	Ambrosini-Spaltro et al. 2011
		SMARCAL1		
Oesophageal	NA	NA	NA	Nakazato et al. 2016
Lymphoma	↓	+	NA	Baradaran-Heravi et al. 2012
Sinal	↓	+	NA	Carroll et al. 2013
		PSPH		
Colorectal	↑	+	NA	Sato et al. 2017
Breast	↑	+	-	Kim et al. 2014
Lung	↑	+	-	Park et al. 2019
Lung	↑	+	NA	Liao et al. 2019

HR: Hazard ratio. ↓: Down-regulation. ↑: Up-regulation. -: Negative effect. +: Positive effect. NA: Not available.

5.4 Discussion

5.4.1 rs9356458 may accelerate CRC tumourigenesis through down-regulation of *MPC1* expression

Ten genes lay in a region of 400kb up- and down-stream of rs9356458, although only *MPC1* was identified for which rs9356458 was an eQTL. Reduced activity of *MPC1*, encoding a mitochondrial pyruvate transporter, has been suggested to play a role in CRC cell growth through altering the maintenance of stem cells (Schell et al., 2014, 2017). Re-expression of *MPC1* repressed the Warburg effect in colorectal cancer cell lines, which is consistent with a causative role in tumourigenesis (Schell et al., 2014). However, it is not currently clear how *MPC1* is regulated or how its functions relate to the known genetic events that cause CRC tumourigenesis (Sandoval et al., 2017). Recently, *MPC1* deficiency has also been shown to accelerate lung adenocarcinoma progression (Zou et al., 2019) and *MPC1* was reported to function as a tumour suppressor and predict the prognosis of patients with renal cell carcinoma (Tang et al., 2019). It has been shown that *MPC1* is down-regulated in numerous human cancers, which also correlates with poor cell survival (Schell et al., 2014). Positive expression of *MPC1* been associated with better OS in prostate cancer patients (Li et al., 2016b) as well as inhibition of proliferation, migration, invasion and stem cell-like properties of gastric cancer cells (Zhou et al., 2017a).

GTEx project data showed that donors homozygous for the alternate allele (A) of rs9356458 were found to have reduced expression of *MPC1* in whole blood. The alternate allele of rs9356458 was associated with an effect size of 0.76, which would indicate that patients with this SNP displayed an increased OS. However, this is contradictory to reports stating that reduced *MPC1* expression correlates with poor cell survival in a number of cancers, including gastric cancer (Zhou et al., 2017a), renal carcinoma (Tang et al., 2019) and colorectal cancer (Schell et al., 2014). Possible reasons for this inconsistency include the fact that the reduced *MPC1* expression from the GTEx project data is in whole blood, as opposed to the tissues analysed in the published reports, and that the gene expression effect size was small.

5.4.2 rs17560791 may affect DNA damage repair, cell cycle progression and maintenance of genome integrity through *SMARCAL1* modulation

Out of nine genes identified within 400 Kb up- and down-stream of rs17560791 (2q35), *SMARCAL1* and *IGFBP2* were identified for which rs17560791 was an eQTL. rs17560791 was an eQTL for *SMARCAL1* in a range of tissues including the oesophagus. The presence of somatic mutations in *SMARCAL1* was reported in oesophageal squamous cell carcinoma, and it is possible that germline mutations impacting the expression of this gene may also play a role in this disease (Nakazato et al., 2016). Furthermore, loss of function alterations in *SMARCAL1* were demonstrated to play a role in the development of childhood lymphoma (Baradaran-Heravi et al., 2012) and carcinoma of the sinus (Carroll et al., 2013).

SMARCAL1 was identified as a critical regulator of the alternative lengthening of telomeres pathway, which cancer cells lacking telomerase activity rely upon to overcome replicative senescence (Cox et al., 2016). SMARCAL1 was also shown to be involved in DNA damage repair, cell cycle progression and maintaining genome integrity, and the development of SMARCAL1 inhibitors as novel anticancer therapies has been suggested (Zhang et al., 2012; Couch et al., 2013; Poole and Cortez, 2017; Puccetti et al., 2019).

rs17560791 was also found to be an eQTL for *AC098820.4*. Expression of *AC098820.4* was found to be dysregulated by rs17560791 in the transverse colon and breast. *AC098820.4* encodes for a long non-coding RNA (lncRNA), which is a natural antisense transcript (NAT) to *SMARCAL1*. NATs are established as important regulators of eukaryotic gene expression (Werner and Swan, 2010). The overexpression of NATs has been associated with poor prognosis in breast cancer (Cayre et al., 2003) as well as epigenetic silencing of tumour suppressor genes (Yu et al., 2008). It has been demonstrated that the dysregulation of lncRNAs is influential in proliferation, angiogenesis, metastasis, invasion, apoptosis, stemness and genomic instability in CRC (Gutschner and Diederichs, 2012; He et al., 2014; Takahashi et al., 2014; Yang et al., 2017; Chen et al., 2017). Recently, lncRNAs have also been shown to play significant roles in breast cancer including the promotion of tumourigenesis (Xu et al., 2017; Rossi et al., 2019) and suppression of metastasis (Kim et al., 2018). LncRNAs represent a potential new paradigm in cancer research that may represent promising therapeutic targets (Bhan et al., 2017; Gutschner et al., 2018).

Donors homozygous for the alternate allele of rs17560791 were also found to show decreased expression of *SMARCAL1*. This is interesting as the effect size associated with this allele (HR 0.77) would appear to confer a protective effect on patients in COIN and COIN-B, although as *SMARCAL1* is involved with DNA damage repair, cell cycle progression and maintaining genome integrity (Zhang et al., 2012; Couch et al., 2013; Poole and Cortez, 2017; Puccetti et al., 2019), it is possible that reduced *SMARCAL1* expression may be associated with inferior survival, although more work is required in this area to gain a clearer understanding of the effect of *SMARCAL1* in terms of CRC prognosis.

5.4.3 rs241477 lies in a gene desert, but may affect gene expression in a trans-regulatory manner

The only gene observed in the local region of rs241477 was the uncharacterised *LOC283585*. rs241477 was not found to have a significant eQTL effect on this gene for any tissues tested in the GTEx Project database. However, SNPs do not necessarily only impact genes in the local region; SNPs can be either ‘cis-regulators’ (eQTLs for genes that are within 4 Mb) or ‘trans-regulators’ (eQTLs for genes that are over 4 Mb away or on a different chromosome (Nicolae et al., 2010). It is therefore possible that rs241477 may have a trans-regulatory eQTL effect on genes outside of the local region in a tissue that was not tested as part of the GTEx Project.

5.4.4 rs4074683 may have a prognostic impact in CRC through altering *PSPH* expression

rs4074683 was identified to be a significant eQTL for the genes *Phosphoserine Phosphatase Pseudogene 1* (*PSPHP1*) and *PSPH* in both the transverse and sigmoid colon, as well as breast and lung tissues. *PSPHP1* is a pseudogene (a section of a chromosome that is an imperfect copy of a functional gene) for *PSPH*. Given the observation that pseudogenes are often found deregulated in cancer progression (Pink et al., 2011), it is possible that *PSPHP1* may play a role in CRC tumourigenesis based on the established links *PSPH* has with the disease (Li et al., 2016a; Sato et al., 2017). *PSPH* has been suggested as a novel prognostic biomarker for CRC due to observed *PSPH* overexpression in CRC tumour tissues and a positive correlation with depth of invasion and distant metastases (Sato et al., 2017). Inhibition of PSPH has also been associated with enhanced efficacy of 5-FU chemotherapy in CRC patients (Li et al., 2016a). In addition, PSPH was recently shown to promote lung cancer progression (Park et al., 2019) and mediate metastasis and proliferation of NSCLC through the EGFR signalling pathway (Liao et al., 2019). In addition, patients with PSPH-positive tumours have been associated with shorter OS in breast cancer (Kim et al., 2014).

Donors homozygous for the alternate allele (T) of rs4074683 showed increased expression of *PSPHP1* in both sigmoid and transverse colon tissues in GTEx project data. rs4074683 appeared to confer a negative prognostic effect on COIN/COIN-B patients in the all wild type subgroup. This correlates with reports that overexpression of *PSPH*, the gene that *PSPHP1* is a pseudogene for, has been observed in CRC tumour tissues and correlates with depth of invasion and distant metastases (Sato et al., 2017). *PSPH*-positive tumours have also been associated with shorter OS in breast cancer patients (Kim et al., 2014) and poor survival in lung cancer patients (Park et al., 2019).

5.4.5 Strengths and limitations of eQTL analyses

Gene expression is strongly altered in tumour tissue compared to normal tissue (Closa et al., 2014), and whilst it is possible that some of the associations identified in healthy tissue still occur in diseased tissue, a current limitation of the eQTL analyses undertaken here is the reference gene expression dataset, which relied on data from the GTEx Project that was collected from normal tissues harvested from healthy donors. However, although this is a limitation of these analyses, the GTEx Project resource was the most comprehensive available – the analysis of eQTL is highly variable across different tissues, because some genes may not be expressed in a specific tissue, thus being undetectable, or because other (epigenetic) regulatory mechanisms of gene expression may interact with the effect of a genetic variation (Nica and Dermitzakis, 2013). It would be interesting to carry out eQTL analyses for the four SNPs identified using a COIN-matched data set that contains gene expression data for cancerous tissues.

5.4.6 Conclusion

The identification of common germline variants that may influence CRC prognosis has the potential to increase our current understanding of CRC pathogenesis and inform clinical management (Manolio, 2010). In this chapter, *in silico* analyses methods conducted to gain an initial insight into the underlying biological mechanisms potentially impacted by four SNPs identified through GWAS analyses in the previous chapter. These potential prognostic biomarkers must now be validated through replication in an independent patient cohort before they can be considered for clinical use (McShane et al., 2005).

Chapter 6

Independent validation of potential germline prognostic biomarkers

6.1 Introduction

The identification of genetic variants that contribute to complex traits is an important challenge in the field of human genetics, and in order to prevent the reporting of chance findings, the validation of any proposed biomarkers is essential (Walther et al., 2009). The gold standard for validation of any genetic study is replication in an additional independent patient cohort (Bush and Moore, 2012) and it is crucial for biomarkers to be robustly validated before they can be considered for implementation in routine patient management (McShane et al., 2005; Van Cutsem et al., 2016).

However, the effects identified through initial GWAS effect size estimation are often subject to an ascertainment bias known as the 'winner's curse', where the actual genetic effect is typically smaller than its estimate (Huang et al., 2018b). This overestimation may cause failure of replication studies due to underestimation of the required validation cohort sample sizes (Zollner and Pritchard, 2007). To account for this phenomenon, replication samples should ideally be larger than the initial GWAS cohort. It is also important for replication studies to be well powered in order to correctly identify any spurious primary GWAS results as false positives (Walther et al., 2009). Initial replication studies should be performed on an independent dataset using patient samples from the same population as the initial GWAS in order to confirm the effect in the target population. Successful replication in the target population can then be followed by replication studies in other populations to determine whether the effect of the SNP is ethnic-specific (Chanock et al., 2007). In terms of CRC, successfully validated prognostic biomarkers have the potential to influence patient management through the introduction of new screening tests to differentiate between specific patient groups (Walther et al., 2009).

This chapter seeks to determine whether the three highly significant variants identified through multivariable GWAS of the full patient cohort can be validated using independent patient cohorts (due to the variant identified in the all wild type subgroup requiring a similarly stratified validation cohort, this variant is not considered in this chapter).

6.1.1 Aims and objectives

- To assess whether the observed association with prognosis of the three variants identified through multivariable GWAS analyses of the full patient cohort could be replicated using an independent series of patients through meta-analyses

6.2 Materials and methods

6.2.1 Patient characteristics

Validation analyses utilised data from 7694 patients across eight independent studies; 357 patients from the Health Professionals Follow-up Study (HPFS), 607 from the Nurses' Health Study (NHS) (Belanger et al., 1978, 1980; Colditz et al., 1997), 323 from the Physician's Health Study (PFS) (Steering Committee of the Physicians' Health Study Research Group, 1989), 270 from the VITamins And Lifestyle Study (VITAL) (White et al., 2004), 1366 from the Women's Health Initiative (WHI) (The Women's Health Initiative Study Group, 1998), 930 from the Quick and Simple and Reliable Trial 2 (QUASAR 2) (Rosmarin et al., 2014), 914 from the Vioxx in Colorectal Cancer Therapy: Definition of Optimal Regimen (VICTOR) study (Pendlebury et al., 2003; Midgley et al., 2010) and 2927 the Short Course Oncology Therapy (SCOT) study (Iveson et al., 2018; Robles-Zurita et al., 2018). Four of these studies (HPFS, NHS, PHS and WHI) are included in the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) (Peters et al., 2012, 2013) and were analysed in the only other known GWAS of CRC prognosis (Phipps et al., 2016). A summary of clinicopathological characteristics of the clinical studies used in validation analyses are shown in Table 6.1.

6.2.1.1 Population-based studies

The HPFS began in 1986, and was conducted as a prospective investigation of dietary aetiologies of heart disease and cancer in men. The purpose of the study was to evaluate a series of hypotheses about men's health relating to nutritional factors to the incidence of serious illnesses such as cancer, heart disease and other vascular diseases. At the beginning of the study, 51,529 male health professionals were enrolled, of which 531 were African-Americans and 877 were Asian-Americans. The HPFS was sponsored by the Harvard School of Public Health and funded by the National Cancer Institue, designed to complement the all-female NHS, which examines similar hypotheses.

The NHS was established in 1976, the original focus of the study was on contraceptive methods, smoking, heart disease and cancer. 238,026 female nurses were sent the original questionnaire, and blood samples were collected from nearly 33,000 participants in 1989-90 to identify potential biomarkers such and hormone levels and genetic markers, followed by a second blood and urine sample from more than 18,700 of the same participants in 2000-02, with DNA collected from an additional 33,000 women in 2001-04 (https://www.nurseshealthstudy.org/about-nhs/history).

The PHS began in 1981 and was a randomised, double-blind, placebo-controlled trial designed to determine whether low-dose aspirin decreases cardiovascular mortality and whether beta carotene reduces the incidence of cancer (Steering Committee of the Physicians' Health Study Research Group, 1989). A total of 22,071 male physicians in the USA between 40 and 84 years of age were randomised to one of four treatment groups; active aspirin and active beta-carotene, active aspirin and beta-carotene placebo, aspirin placebo and active beta-carotene,

or both placebos (Manson et al., 1991). The results of the beta carotene component of the study found neither a benefit or harm in terms of malignant neoplasms, cardiovascular disease, or death from all causes (Hennekens et al., 1996).

The VITAL study was a cohort study of the associations of supplement use and cancer risk comprising of a total of 77,738 men and women between 50-76 years old, who entered the study between 2000-02 (White et al., 2004). Participants completed a detailed questionnaire on supplement use, diet and other cancer risk factors, and 70% provided DNA through self-collected buccal cell specimens (White et al., 2004). The investigators reported that four supplements were significantly associated with risk factors for breast, prostate, lung and colorectal cancers (White et al., 2004).

The WHI was a long-term national health study that focused on strategies for preventing the major causes of death, disability and frailty in older women; specifically heart disease, osteoporotic fractures and cancer. The WHI originally enrolled 161,808 women aged 50-79 between 1993-1998 and had two major parts; a clinical trial (n=68,132) and an observational study (n=93,676). The clinical trial tested three prevention strategies; hormone therapy trials, dietary modification trial and a calcium/vitamin D trial, while the observational study examined the relationship between lifestyle, health and risk factors and disease outcomes.

6.2.1.2 Clinical trials

The QUASAR 2 study was an open-label, randomised controlled trial consisting of 1952 eligible patients (1941 with assessable data) enrolled between April 2005 and October 2010 (Kerr et al., 2016). Patients from 170 hospitals across seven countries and over 18 years of age with WHO PS scores of 0 or 1 who had undergone potentially curative surgery for histologically proven Stage III or high-risk Stage II CRC were randomly assigned to receive eight three-week cycles of oral capecitabine alone, or oral capecitabine plus 16 cycles of intravenous bevacizumab on day one of each cycle, and the primary endpoint was disease-free survival (Kerr et al., 2016). The results found that the addition of bevacizumab to capecitabine in the adjuvant setting for CRC conferred no benefit to patients, and it therefore should not be used (Kerr et al., 2016).

The VICTOR trial was a double-blind randomised trial assessing whether the COX-2 inhibitor rofecoxib could reduce recurrence and improve survival when administered in the adjuvant setting of CRC (Midgley et al., 2010). 2434 patients who had undergone potentially curative surgery and completion of adjuvant therapy for Stage II and III CRC were entered into the study and randomly assigned to receive rofecoxib or placebo. The primary endpoint was OS, and where FFPE tumour tissue samples were available, COX-2 expression was evaluated by immunohistochemistry and correlated with clinical outcome (Midgley et al., 2010). The trial was terminated early due to the worldwide withdrawal of rofecoxib, at which time 1167 patients had received rofecoxib and 1160 patients had received placebo, for median treatment durations of 7.4 and 8.2 months, respectively. The results of the trial found no difference in OS between the two groups (HR 0.97, 95% CI 0.81-1.16, P=0.75); tumour COX-2 expression was assessed for 871 patients, but no prognostic or predictive effects were observed (Midgley et al., 2010).

The SCOT trial was an international, randomised, phase 3, non-inferiority trial which assessed the efficacy, toxicity and cost-effectiveness of three vs. the standard six months of adjuvant chemotherapy in CRC (Iveson et al., 2018; Robles-Zurita et al., 2018). In total, 6088 patients aged 18 or older with high-risk Stage II and Stage III CRC from 244 centres were randomly assigned to receive three or six months of adjuvant oxaliplatin-containing chemotherapy (either XELOX or OxMdG), and the primary endpoint of the study was disease-free survival (Iveson et al., 2018; Robles-Zurita et al., 2018). The results of the study were that three months of oxaliplatin-containing chemotherapy was non-inferior to six months of the same therapy, suggesting that the shorter duration confers similar survival outcomes with a better quality of life, and may represent a new standard of care (Iveson et al., 2018; Robles-Zurita et al., 2018).

6.2.2 Study design and statistical analysis methods

6.2.2.1 Selection of SNPs for validation analyses

Three of the four SNPs for which functional analyses were performed (rs9356458 at 6q27, rs17560791 at 2q35 and rs241477 at 14q31.3) were considered for validation through meta-analyses. These SNPs were all identified through GWAS analyses of the full patient cohort, whereas rs4074683 (7p11.2) was identified in GWAS analyses of the all wild type subgroup. The validation cohorts were not stratified for underlying somatic prognostic factors, and therefore this SNP was not considered in these analyses.

6.2.2.2 Power calculations

Statistical power calculations for the combined validation cohort were performed using the online calculator at http://www.sample-size.net/sample-size-survival-analysis/.

6.2.2.3 Meta-analyses

All meta-analyses were performed using the meta package in R. The metagen function was used to perform inverse-variance weighted meta-analysis, as this method is associated with a lower standard error than other methods and is generally preferred (de Bakker et al., 2008). Analyses were stratified firstly by study, and secondly by using the studies with relevant data to form a subset of mCRC patients only, in order to match the clinicopathological stage of patients in the combined COIN and COIN-B cohort. All meta-analyses used multivariable data. The HPFS, NHS, PHS, VITAL and WHI studies were adjusted for age at diagnosis, sex, GWAS batch and the first 13 principal components. The QUASAR 2, VICTOR and SCOT studies were adjusted for age, sex, location of the primary tumour and AJCC Stage at diagnosis.

The Cochrans' Q and I^2 tests of heterogeneity were performed in order to ascertain whether a fixed- or random-effects model of meta-analysis would be employed for each SNP. No heterogeneity was present for rs9356458 and rs17560791, therefore these SNPs were analysed under a fixed effects model. Moderate heterogeneity was found for rs241477, therefore a random-effects model was applied for meta-analysis of this SNP.

Forest plots were created in order to visualise the results of the study-stratified meta-analyses, highlighting the HRs and 95% CIs of each study and creating a combined HR and 95% CI for all studies. Funnel plots were also created in order to check for the presence of any underlying reporting bias (Sterne et al., 2011). Forest plots and funnel plots were created using the forest and funnel functions, respectively.

6.3 Results

6.3.1 Validation cohorts

Survival data was available from five population-based studies; 357 patients (209 events) in HFPS, 607 patients (258 events) in NHS, 323 patients (199 events) in PHS, 270 patients (109 events) in VITAL, 1366 patients (410 events) in WHI and from three clinical trials; 930 patients (166 events) in QUASAR 2, 914 patients (107 events) in VICTOR and 2927 patients (186 events) in SCOT were used in validation analyses (Table 6.1).

Table 6.1: Clinicopathological data for validation cohorts.

Study	HPFS	NHS	PHS	VITAL	WHI	QUASAR 2	VICTOR	SCOT
Number of patients	357	607	323	270	1366	930	914	2927
Number of events	209	258	199	109	410	166	107	186
Age								
Mean (SD)	72 (8.7)	69 (8.6)	71 (9.6)	70 (6.5)	72 (7.2)	64 (9.8)	64 (9.5)	64 (9.0)
<20	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
20-49	2 (0.6)	11 (1.8)	5 (1.5)	0 (0.0)	0 (0.0)	79 (8.5)	66 (7.2)	206 (7.0)
50-59	28 (7.8)	64 (10.5)	36 (11.1)	23 (8.5)	61 (4.5)	203 (21.8)	244 (26.7)	581 (19.8)
60-69	101 (28.3)	227 (37.4)	110 (34.1)	91 (33.7)	461 (33.7)	375 (40.3)	340 (37.2)	1363 (46.6)
70-79	152 (42.6)	236 (38.9)	105 (32.5)	144 (53.3)	644 (47.1)	250 (26.9)	243 (26.6)	740 (25.3)
80-89	74 (20.7)	69 (11.4)	67 (20.7)	12 (4.4)	200 (14.6)	23 (2.5)	21 (2.3)	37 (1.3)
Male	357 (100)	0 (0.0)	323 (100)	146 (54)	0 (0.0)	534 (57.4)	592 (64.8)	1793 (61.3)
Female	0 (0.0)	607 (100.0)	0 (0.0)	124 (46)	1366 (100.0)	396 (42.6)	322 (35.2)	1134 (38.7)
AJCC Stage								
I	47 (13.2)	142 (23.4)	64 (19.8)	72 (26.6)	421 (30.8)	0 (0.0)	0 (0)	0 (0.0)
II	29 (8.1)	168 (27.7)	62 (19.2)	48 (17.7)	377 (27.6)	336 (36.1)	468 (51.2)	584 (20.0)
III	81 (22.7)	148 (41.5)	59 (18.3)	47 (17.4)	332 (24.3)	594 (63.9)	446 (48.8)	2343 (80.0)
IV	23 (6.4)	81 (13.3)	44 (13.6)	31 (11.5)	178 (13.0)	0 (0.0)	0 (0.0)	0 (0.0)

Uncharacterised mutants and failed genotyping data omitted. Percentages are shown in parentheses unless otherwise stated. SD: Standard deviation. NA: Not assessed. Some percentages do not add up to 100% due to patients without cancer involved in the studies.

6.3.2 Power calculations

The statistical power associated with the variants rs9356458, rs17560791 and rs241477 in the combined validation cohort using their known MAFs and HRs from multivariable GWAS analsyes was 98%, 97% and 82%, respectively (Table 6.2).

Table 6.2: Statistical power to detect associations with OS in validation analyses.

Variant	Cytoband	N	Events	Ref	Alt	MAF	HR	Power
rs9356458	6q27	7694	1644	G	A	0.41	0.76	0.98
rs17560791	2q35	7694	1644	G	C	0.49	0.77	0.97
rs241477	14q31.3	7694	1644	T	A	0.12	0.71	0.82

N: Sample size. Ref: Reference allele. Alt: Alternate allele (allele analysed). MAF: Minor allele frequency. HR: Hazard ratio.

6.3.3 Meta-analyses

All meta-analyses used multivariable data. The HPFS, NHS, PHS, VITAL and WHI studies were adjusted for age at diagnosis, sex, GWAS batch and the first 13 principal components. The QUASAR 2, VICTOR and SCOT studies were adjusted for age, sex, location of the primary tumour and AJCC Stage at diagnosis.

6.3.3.1 rs9356458

The individual results of all studies for rs9356458 were as follows; HPFS: HR 0.91, 95% CI 0.73-1.13, *P*=0.41; NHS: HR 1.00, 95% CI 0.83-1.19, *P*=0.99; PHS: HR 1.04, 95% CI 0.85-1.26, *P*=0.73; VITAL: HR 0.88, 95% CI 0.65-1.17, *P*=0.37; WHI: HR 0.97, 95% CI 0.84-1.12, *P*=0.68; QUASAR 2: HR 1.18, 95% CI 0.95-1.46, *P*=0.14; VICTOR: HR 0.88, 95% CI 0.66-1.18, *P*=0.39; SCOT: HR 1.05, 95% CI 0.85-1.30, *P*=0.63. The results of meta-analyses for rs9356458 in a total of 7694 patients (1644 deaths) were not significant (HR 0.99, 95% CI 0.93-1.07, *P*=0.88; Figure 6.1, Table 6.3).

Table 6.3: Meta-analyses results for the independent validation of rs9356458.

Study	N	Events	HR	95% CI	*P*
HPFS	357	209	0.91	0.73-1.13	0.41
NHS	607	258	1.00	0.83-1.19	0.99
PHS	323	199	1.04	0.85-1.26	0.73
VITAL	270	109	0.88	0.65-1.17	0.37
WHI	1366	410	0.97	0.84-1.12	0.68
QUASAR 2	930	166	1.18	0.95-1.46	0.14
VICTOR	914	107	0.88	0.66-1.18	0.39
SCOT	2927	186	1.05	0.85-1.30	0.63
TOTAL	7694	1644	0.99	0.93-1.07	0.88

All studies used overall survival as the primary outcome measure. Meta-analysis of rs9356458 was performed using a fixed effects model as no heterogeneity was present (I^2=0.00%), Q=4.80, P^{HET}=0.70. N: Sample size. HR: Hazard ratio. CI: Confidence interval. *P*: P-value. I^2: I^2 Test value. Q: Cochran's Q Test value. P^{HET}: P-value of heterogeneity.

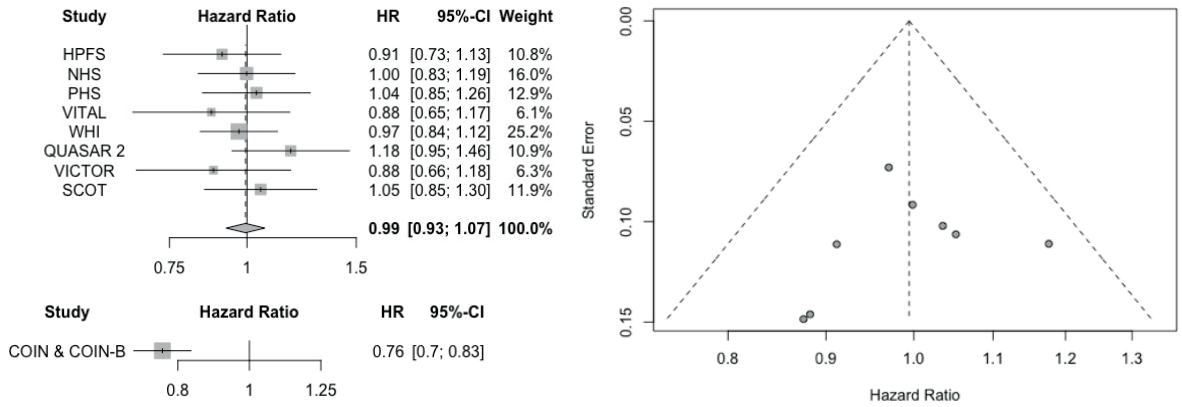

Figure 6.1: Meta-analyses results for the independent validation of rs9356458. HR: Hazard ratio. CI: Confidence interval.

6.3.3.2 rs17560791

The individual results of all studies for rs17560791 were as follows; HPFS: HR 1.09, 95% CI 0.89-1.33, *P*=0.41; NHS: HR 0.88, 95% CI 0.73-1.06, *P*=0.10; PHS: HR 1.01, 95% CI 0.81-1.25, *P*=0.96; VITAL: HR 0.89, 95% CI 0.67-1.19, *P*=0.43; WHI: HR 0.89, 95% CI 0.77-1.04, *P*=0.13; QUASAR 2: HR 0.94, 95% CI 0.75-1.19, *P*=0.61; VICTOR: HR 1.03, 95% CI 0.78-1.37, *P*=0.83; SCOT: HR 0.99, 95% CI 0.80-1.22, *P*=0.92. Meta-analyses of rs17560791 were also not significant (HR 0.95, 95% CI 0.89-1.03, *P*=0.20; Figure 6.2, Table 6.4).

Table 6.4: Meta-analyses results for the independent validation of rs17560791.

Study	N	Events	HR	95% CI	*P*
HPFS	357	209	1.09	0.89-1.33	0.41
NHS	607	258	0.88	0.73-1.06	0.10
PHS	323	199	1.01	0.81-1.25	0.96
VITAL	270	109	0.89	0.67-1.19	0.43
WHI	1366	410	0.89	0.77-1.04	0.13
QUASAR 2	930	166	0.94	0.75-1.19	0.61
VICTOR	914	107	1.03	0.78-1.37	0.83
SCOT	2927	186	0.99	0.80-1.22	0.92
TOTAL	7694	1644	0.95	0.89-1.03	0.20

All studies used overall survival as the primary outcome measure. Meta-analysis of rs17560791 was performed using a fixed effects model as no heterogeneity was present (I^2=0.00%), Q=4.00, P^{HET}=0.80. N: Sample size. HR: Hazard ratio. CI: Confidence interval. *P*: P-value. I^2: I^2 Test value. Q: Cochran's Q Test value. P^{HET}: P-value of heterogeneity.

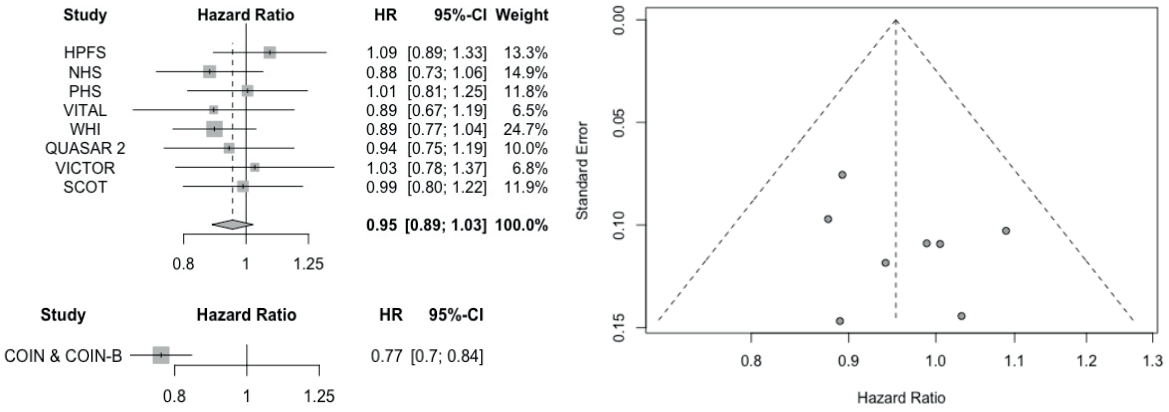

Figure 6.2: Meta-analyses results for the independent validation of rs17560791.
HR: Hazard ratio. CI: Confidence interval.

6.3.3.3 rs241477

The individual results of all studies for rs241477 were as follows; HPFS: HR 0.72, 95% CI 0.52-0.99, *P*=0.05; NHS: HR 1.07, 95% CI 0.82-1.39, *P*=0.14; PHS: HR 1.69, 95% CI 1.14-2.52, *P*=9.5x10^{-3}; VITAL: HR 1.19, 95% CI 0.76-1.88, *P*=0.45; WHI: HR 1.04, 95% CI 0.83-1.30, *P*=0.73; QUASAR 2: HR 0.74, 95% CI 0.55-0.99, *P*=0.04; VICTOR: HR 1.08, 95% CI 0.71-1.64, *P*=0.73; SCOT: HR 1.09, 95% CI 0.79-1.50, *P*=0.61. Meta-analyses results for rs241477 were not significant (HR 1.02, 95% CI 0.86-1.21, *P*=0.82; Table 6.5, Figure 6.1C).

Table 6.5: Meta-analyses results for the independent validation of rs241477.

Study	N	Events	HR	95% CI	*P*
HPFS	357	209	0.72	0.52-0.99	0.05
NHS	607	258	1.07	0.82-1.39	0.14
PHS	323	199	1.69	1.14-2.52	9.5x10^{-3}
VITAL	270	109	1.19	0.76-1.88	0.45
WHI	1366	410	1.04	0.83-1.30	0.73
QUASAR 2	930	166	0.74	0.55-0.99	0.04
VICTOR	914	107	1.08	0.71-1.64	0.73
SCOT	2927	186	1.09	0.79-1.50	0.61
TOTAL	7694	1644	1.02	0.86-1.21	0.82

All studies used overall survival as the primary outcome measure. Meta-analysis of rs241477 was performed using a random effects model due to the presence of heterogeneity (I^2=53.1%), Q=14.9, P^{HET}=0.04. N: Sample size. HR: Hazard ratio. CI: Confidence interval. *P*: P-value. I^2: I^2 Test value. Q: Cochran's Q Test value. P^{HET}: P-value of heterogeneity.

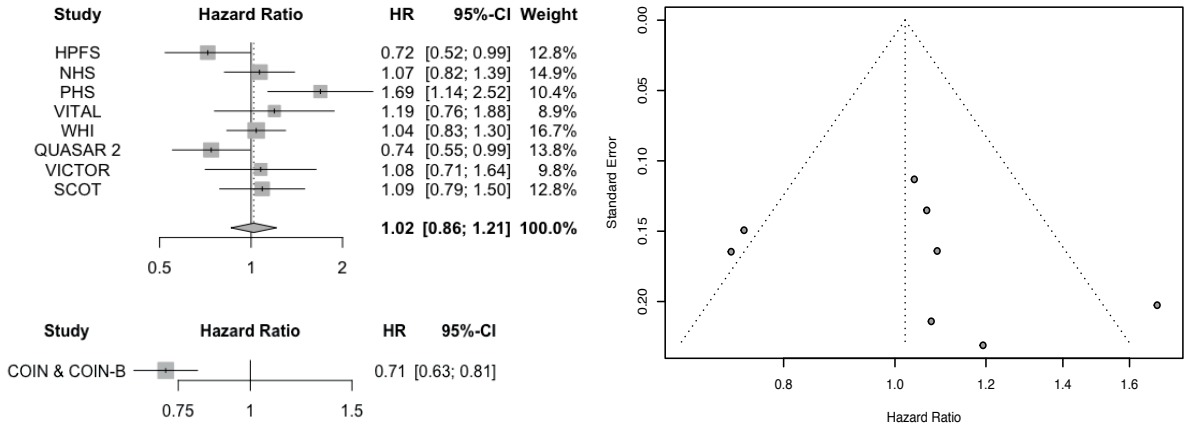

Figure 6.3: Meta-analyses results for the independent validation of rs241477.
HR: Hazard ratio. CI: Confidence interval.

6.4 Discussion

6.4.1 Possible reasons for the lack of replication of potential prognostic biomarkers

It is essential that variants identified as potential prognostic biomarkers are validated in an independent cohort of patients prior to being considered for clinical application (McShane et al., 2005; Van Cutsem et al., 2016). In this chapter, meta-analyses were performed using survival data from three SNPs identified as significantly associated with prognosis through multivariable GWAS analyses (rs9356458 [6q27], rs17560791 [2q35] and rs241477 [14q31.3]) to determine whether similar effects could be observed in an independent patient series.

Validation analyses had $>80\%$ power to detect an association with survival for rs241477 and $>90\%$ power to detect associations with survival for rs9356458 and rs17560791 based upon the HRs previously obtained for these variants under multivariable analyses (0.71, 0.76 and 0.77, respectively). However, due to winner's curse, it should be assumed that the true effect size is considerably lower than the estimated value (Garner, 2007). To account for this, it has been suggested that in order to replicate the most significant original associations from a GWAS, the sample size of the validation cohort may need to be twice the size of the original study (Garner, 2007). While the sample size of the combined validation cohort analysed here satisfied this criterion (n=7694), due to the large number of patients included in validation analyses having earlier stages of cancer, the number of events was significantly smaller (n=1644). The number of events for rs9356458, rs17560791 and rs242477 were n=1139, n=932 and n=1122, respectively, hence the number of events used in the validation cohort may not be large enough to account for the potential effects of winner's curse in the original GWAS (Garner, 2007). Using an arbitrarily smaller HR of 0.85 as an example, the power available to detect each variant reduced drastically to 19% for rs9356458, 21% for rs17560791 and negligible power for rs241477. Therefore, if the HRs from the original GWAS were overinflated due to winner's curse, it is possible that this may be a reason why the associations between these variants and survival have not been replicated here.

The lack of replication may also be due to differences in patient characteristics between the combined COIN and COIN-B cohort and the validation cohorts. The majority of patients in the validation cohorts had Stage II/III disease, with only four of the eight studies analysed containing patients with Stage IV disease; HPFS (23/357 [6.4%]), NHS (81/607 [13.3%]), PHS (44/323 [13.6%]) and WHI (178/1366 [13.0%]). There are well-established differences between early and late-stage CRC patients that may have confounded the results, including the significant difference in survival between MSI-positive patients, with MSI conferring an improved prognosis in Stage II patients compared to poor survival in Stage IV patients (Popat et al., 2005; Smith et al., 2013). There is also a possibility that the associations identified in the initial GWAS are mCRC-specific, but due to the small numbers of mCRC patients in the validation cohort, there was insufficient power to perform mCRC-specific validation analyses.

The combined validation cohort contained patients from American studies (HPFS, NHS, PHS, VITAL and WHI), albeit all patients included self-reported they were of European descent (Phipps et al., 2016). Even so, there are likely to be underlying differences in genetic architecture between the primary and follow-up cohorts in these analyses. LD between markers can differ greatly between populations (Slatkin, 2008; Manolio et al., 2009), which might have had a confounding effect on the results obtained for the three SNPs between the cohorts.

Another limitation of the validation cohorts is the number of prognostic covariates included in their analysis models. As the SNPs identified in the combined COIN and COIN-B cohort were found through multivariable analyses that included a number of prognostic factors, a potential reason for these SNPs not being significant in these validation cohorts is their limited clinical information. It is possible that with a large enough dataset that additionally provides meta data on a larger number of covariates matching those of the combined COIN and COIN-B cohort, a significant effect could be seen for one or more of these variants.

6.4.2 Conclusion

In this chapter, three SNPs from GWAS analyses of the full patient cohort were analysed in meta-analyses using data from independent patient cohorts in order to ascertain whether these variants could be validated as prognostic biomarkers for CRC. None of the SNPs identified through GWAS analyses were validated in the independent validation cohorts used. However, this could be due to limitations in these validation cohorts in terms of sample size for mCRC patients and lack of prognostic covariates. Further work is required with validation cohorts more similarly matched to the GWAS cohort if possible, which may lead to a higher likelihood of validating these variants as prognostic markers for mCRC.

Chapter 7

General discussion

7.1 Novel and confirmatory findings from this work

7.1.1 Somatic mutations, MSI and survival

In agreement with previous studies, the work in this thesis has identified that mutations in *KRAS*, *BRAF* and *NRAS* and MSI-positive tumours are associated with a poor prognosis in mCRC (Richman et al., 2009; Tran et al., 2011; Eklof et al., 2013; Phipps et al., 2013; Smith et al., 2013; Schirripa et al., 2015; Wang et al., 2018). Statistically significant observations were also identified when analysing mutations in terms of their respective codons. Mutations in codons 12 and 13 of *KRAS* were found to confer a poor prognosis compared to *KRAS* wild type patients, which is in line with earlier reports (Imamura et al., 2012; Phipps et al., 2013). Although patients with mutations in codon 61 of *KRAS* appeared to have an inferior prognosis to those with *KRAS* wild type tumours, this observation was not statistically significant. Mutations in codon 61 of *KRAS* have previously been observed as not being significantly associated with clinical outcome in CRC (Imamura et al., 2014). This may be due to the relatively rare occurrence of codon 61 mutations compared with other mutations in *KRAS*, meaning studies often have a low sample size for these mutations, and subsequently low statistical power with which to detect prognostic effects.

BRAF c.1799T $>$ A (p.V600E) mutations are well established as conferring a poor prognosis in mCRC (Richman et al., 2009; Tran et al., 2011; Van Cutsem et al., 2011), which supports the findings reported in this study. However, less is known about the prognostic impact of *BRAF* c.1781A $>$ G (p.D594G) mutations. Here, patients with *BRAF* c.1781A $>$ G (p.D594G) mutant tumours did not display a significant difference in OS to those with *BRAF* wild type tumours. This may potentially be due to the small sample size of patients with *BRAF* c.1781A $>$ G (p.D594G) mutant tumours in this cohort and the associated lack of power. Further investigation with a larger cohort is needed to determine the prognostic effects of this mutation in mCRC.

Mutations in codon 61 of *NRAS* conferred an inferior prognosis to *NRAS* wild type tumours. This observation is in agreement with another report identifying *NRAS* mutations as an indicator of poor prognosis when analysed as a whole (Schirripa et al., 2015). However, perhaps due to their rarity in CRC (Irahara et al., 2010), there is currently very little in the literature analysing the effects of mutations in individual codons of *NRAS* on prognosis. While *NRAS* codon 61 mutant tumours conferred an inferior prognosis to *NRAS* tumours here, the picture is less clear for *NRAS* codon 12 and 13 mutant tumours. No significant difference in OS was observed between patients with *NRAS* codon 12 or 13 mutant tumours when compared to *NRAS* wild type tumours. This may be due to the small sample size of patients with *NRAS* codon 12 and 13

mutant tumours and the associated lack of power to detect prognostic effects for this subgroup of *NRAS* mutant tumours, and warrants further investigation using a larger cohort.

MSI-positive tumours conferred a poor prognosis, in agreement with other studies in mCRC (Tran et al., 2011; Smith et al., 2013). It is important to note that MSI confers a good prognosis in early-stage CRC (Popat et al., 2005; Bertagnolli et al., 2009; Hutchins et al., 2011).

7.1.2 Germline mutations and survival

To date, only two germline variants have been robustly associated with CRC prognosis (Smith et al., 2015; Phipps et al., 2016). Two novel SNPs associated with survival to mCRC at a level of genome-wide significance were identified through the work undertaken for this thesis. A further two SNPs approaching genome-wide significance were also identified, one through GWAS analyses of the full patient cohort and one from a GWAS of a subgroup of patients who were wild type for somatic mutations in the *KRAS*, *BRAF*, *NRAS* oncogenes, and were MSS. These SNPs were associated with a longer OS, which suggests their minor alleles may have a protective effect on the patients that carry them. According to the CD/CV hypothesis, it is likely that these SNPs could play a role in prognosis in conjunction with other germline variants (Bush and Moore, 2012).

The three highly significant SNPs from the full patient cohort were taken forward to validation analyses. Although none of these SNPs were validated in the independent cohort, it is a possibility that this may be due to differences in patient characteristics between the COIN and COIN-B cohort and the validation cohort. However, even though the Bonferroni correction used to limit the amount of false-positive associations is conservative, the possibility exists that these associations are no more than artefacts.

7.2 Clinical utility of this work

7.2.1 Somatic mutations and MSI

The role of somatic mutations as prognostic biomarkers for mCRC in the clinical setting is currently limited, with only tumour *BRAF* mutation testing currently routinely assessed due to it being a strong negative prognostic factor (Van Cutsem et al., 2014, 2016). From a clinical perspective, almost all current knowledge of *BRAF* mutant CRC has been derived from patients with *BRAF* c.1799T$>$A (p.V600E) mutant tumours (Jones et al., 2017). Although some groups have begun to investigate the clinical characteristics of patients with non-V600E mutations (Cremolini et al., 2015), due to their rarity, study sample sizes are small and the power associated with these studies is low. It has therefore been suggested that it is difficult to draw robust conclusions from this report alone (Jones et al., 2017).

Here, *BRAF* mutations conferred a median reduction in OS of 295 days compared to *BRAF* wild type tumours. These results are in agreement with published data showing a clear prognostic impact for *BRAF* mutations (Farina-Sarasqueta et al., 2010; Tran et al., 2011; Lochhead

et al., 2013), and also support other findings that a difference in prognosis exists between V600E and non-V600E mutations in *BRAF* (Cremolini et al., 2015; Jones et al., 2017). This is a significant observation, as although the vast majority of *BRAF* mutations are V600E, differences in prognosis may mean that functional effects exerted by individual *BRAF* mutants differ, which subsequently may have an impact on the *BRAF* inhibitors chosen for patients. Targeted *BRAF* V600E inhibitors such as vemurafenib that significantly improved OS and PFS in *BRAF* V600E mutated metastatic melanoma have so far failed to achieve similar results in *BRAF* V600E mutated mCRC; BRAF inhibitor monotherapy resulted in fewer than 10% of responders and a maximum PFS of 4.3 months (Kopetz et al., 2015; Hyman et al., 2015). It is indicated that for patients with *BRAF* mutant mCRCs, combination strategies might be more promising and further work is required in this area (Taieb et al., 2019), an example of which is the recent BEACON clincial trial, which tested the efficacy of the *BRAF* inhibitors encorafenib and binimetinib in combination with cetuximab in patients with *BRAF* mutant mCRC (Kopetz et al., 2019). The results of the trial showed that a combination of encorafenib, binimetinib and cetuximab resulted in significantly longer OS and a higher response rate than standard therapy (Kopetz et al., 2019). It is also unclear whether CRCs with non-V600E *BRAF* mutations exhibit the same resistance to anti-EGFR therapy as cancers with *BRAF* V600E mutations (Jones et al., 2017). The clinical impact of non-V600E *BRAF* mutations is currently unknown, and non-V600E mutations are an area of active research (Cremolini et al., 2015).

KRAS mutational status is a negative predictive biomarker for therapeutic choices involving EGFR antibody therapies, although the prognostic role of *KRAS* mutations remains unclear (Imamura et al., 2012). A number of reports are conflicting (Barault et al., 2008; Ogino et al., 2009; Zlobec et al., 2010; Farina-Sarasqueta et al., 2010). Consequently, current clinical guidelines do not advise tumour *KRAS* mutation testing for prognostic assessment (Van Cutsem et al., 2016). In this report, *KRAS* mutations conferred a median reduction in OS of 131 days compared to *KRAS* wild type tumours. Until recently, mutant *KRAS* has been considered an undruggable target (Malumbres and Barbacid, 2003; Roberts and Stinchcombe, 2013; Cox et al., 2014; Porru et al., 2018), although studies are continuing to develop new approaches for blocking KRAS activity (Hobbs et al., 2016b), with recent studies confirming *KRAS* inhibitors have progressed into clinical development (Mullard, 2019).

NRAS mutations are now routinely included in clinical testing alongside testing *KRAS* mutations prior to EGFR inhibitor therapy for mCRC, although the clinical implications of *NRAS* mutations beyond lack of response to anti-EGFR therapy remains unknown (Cercek et al., 2017). Here, *NRAS* mutations conferred a median reduction in OS of 112 days compared to *NRAS* wild type tumours, in agreement with other reports (Schirripa et al., 2015; Cercek et al., 2017).

MSI status is used in the clinical setting for early-stage cancer due to the positive prognostic impact MSI confers on Stage II CRC patients (Labianca et al., 2013). Recently, immunotherapy has emerged as the fourth pillar of cancer treatment, alongside surgery, radiation and

chemotherapy (Iwai et al., 2017). It has been shown that mCRC patients with MSI-positive tumours are susceptible to immune checkpoint inhibitors, such as Programmed Cell Death Protein 1 (PD-1) and Programmed Cell Death Ligand 1 (PD-L1) inhibitors (Oliveira et al., 2019). Here, patients with MSI-positive tumours had a median reduction in survival of 244 days compared to those with MSS tumours, in agreement with other studies of patients with mCRC (Tran et al., 2011; Smith et al., 2013).

7.2.2 Germline variants

To date, no germline prognostic biomarkers for mCRC have made it into the clinic (Van Cutsem et al., 2016). Due to germline variants in complex diseases conferring modest effect sizes (as per the CD/CV hypothesis) (Frazer et al., 2009; Bush and Moore, 2012), it is unlikely that the effect size for a single SNP associated with prognosis will be clinically actionable when taken in isolation. Therefore, it may be possible to combine these effect sizes with those of other germline variants and somatic prognostic factors to create effect sizes that are clinically actionable.

Even though no germline variants are currently routinely in clinical use for sporadic CRC, they present a promising area for future research, which may improve the accuracy of prognosis and subsetting patients for treatment strategies. Recently, germline variants have been associated with severe toxicities during chemotherapy, and cancer-associated variants are associated with clinical outcome in Stage III CRC patients, with patients carrying cancer-associated variants having better disease-free survival compared to those who did not (Lin et al., 2019). Further epidemiological and functional investigations of germline variants may add to our understanding of CRC pathogenesis, and may ultimately lead to personalised strategies for the treatment of CRC (Zhang et al., 2014).

7.3 Strengths and limitations of this work

Combined, the clinical trials COIN and COIN-B represent the largest study regarding the efficacy of moAB therapy in mCRC to date. Cetuximab treatment did not significantly influence the OS of patients (Maughan et al., 2011), which was confirmed through the analyses reported in this study. This finding, coupled with the homogeneous nature of COIN and COIN-B, enabled the two cohorts to be combined, which increased the statistical power of the subsequent analyses in this thesis. However, a larger sample size may have further increased the power of these analyses, enabling the detection of of smaller effect sizes and further limiting the number of false positive results (Hirschhorn and Daly, 2005; Biau et al., 2008; Button et al., 2013). While the sample size of this study is a limitation, is it one that is common to all studies of this nature (Phipps et al., 2016).

In an attempt to limit the number of false positive results reported from GWASs, the genome-wide level of significance ($P < 5.0\text{x}10^{-8}$) was proposed as an equivalent to $P < 0.05$ after a Bonferroni correction for one million independent tests (Risch and Merikangas, 1996). While this stringent threshold reduces the chance of false positives, there is subsequently a high likelihood of false negative findings; a limitation inherent to the GWAS approach (Phipps et al., 2016). This method is often considered punitively conservative and a potential over-correction (Sham and Purcell, 2014; Fadista et al., 2016), due to the violation of the assumption of independence between tests by SNPs in LD with each other (Hirschhorn and Daly, 2005). However, the Bonferroni method is established as the *de facto* standard multiple testing method for GWAS analyses (Ball, 2013; Sham and Purcell, 2014; Fadista et al., 2016), and due to the increased computational power associated with other methods such as permutation testing (Hirschhorn and Daly, 2005), the Bonferroni correction was implemented here.

Based on statistical power calculations, only variants with MAF ≥ 0.05 were included in GWAS analyses. Greater statistical power is often associated with common variants (MAF ≥ 0.05), and in accordance with the CD/CV hypothesis, the majority of significant GWAS findings involve common SNPs (Bush and Moore, 2012). However, although a threshold of MAF ≥ 0.05 has been used by others in prognostic GWAS analyses for CRC (Phipps et al., 2016), it is possible that the implementation of this threshold may cause some variants with lower allele frequencies and a genuine association with CRC prognosis to have been missed. However, minimising false positives in GWAS analyses is essential, and due to the conservative nature of the Bonferroni correction, associations identified between a variant and phenotype at a level of genome-wide signficance can be interpreted as robust (Hirschhorn and Daly, 2005).

A benefit of analysing the COIN and COIN-B trial data over other studies was the vast amount of clinical molecular data that had been collected from patients, which enabled multivariable survival analyses to be performed. This is beneficial in clinical studies in order to assess prognosis with respect to several contributing factors simultaneously, and offers estimates of the

strength of effect for each constituent factor (Bradburn et al., 2003). Effect sizes obtained through multivariable analyses may therefore be more informative than those from univariable analyses (Bradburn et al., 2003), altough the lack of similar clinicopathological data in the validation cohorts may have hindered attempts to replicate the variants identified as prospective prognostic biomarkers.

The associations with survival demonstrated by the three SNPs identified through GWAS analyses did not replicate in the meta-analyses. Currently, these SNPs are not currently validated prognostic biomarkers for mCRC. However, reasons for this lack of validation may be due to limitations in the validation cohort, such as lack of Stage IV patient data and insufficient number of prognostic covariates.

Although the clinical molecular data collected for the COIN and COIN-B trial is a key strength of the work in this thesis, there are limitations in relation to the genotyping methods used that might limit the accuracy of some of the findings of this study. Mutations in *KRAS*, *BRAF* and *NRAS* were tested for using a combination of pyrosequencing and Sequenom, which tested for mutations in hotspots across the gene, but not the entire gene.

The years since the COIN and COIN-B dataset was genotyped has seen the advent of new methods of mutational testing such as NGS, which often identifies mutations with unclear clinical or prognostic implications (Jones et al., 2017). Advances in NGS have allowed for more comprehensive testing of multiple mutational hotspots within various genes of interest (Jones et al., 2017). It has been shown that while there is a high concordance rate between NGS and standard testing for *KRAS*, NGS revealed mutations that are not tested for with standard *KRAS* assays, which may have a clinical impact (Kothari et al., 2014). A more recent study identified 20% of patients who were originally classified as *KRAS* wild type mCRC were subsequently found to have mutations in exon 3 or 4 of *KRAS* or *NRAS* (Allegra et al., 2016). It is clear that the sequencing methods used in this study have their limitations, especially when considered against more recent alternatives such as NGS, but as all wet lab work was carried out prior to the commencement of this project, it is a limitation that was outside of the candidate's control.

7.4 Future work

The main focus of future work resulting from this thesis would be to attempt replication of the three highly significant germline variants in more suitable validation cohorts. If possible, validation cohorts that are stage-matched with COIN and COIN-B and have similar clinicopathological patient data could result in these variants being robustly validated as prognostic biomarkers for mCRC. The sample size of the cohort would also need to be larger in order to account for the effect of winner's curse that is likely to be a source of bias in the effect sizes reported in this study (Oetting et al., 2017).

Although the clinical molecular data collected for the COIN trial is extensive, the sequencing methods used are now dated. As mentioned in the previous section, pyrosequencing, Sequenom and Sanger sequencing have limitations that could be minimised through the use of NGS, as this method has now mostly superseded conventional molecular biology methodologies (Behjati and Tarpey, 2013). It would be interesting to reanalyse the COIN and COIN-B cohorts after having the samples resequenced using NGS, as this would be likely to increase the accuracy of the genotype calling, ensuring the 'all wild type' subgroup really is wild type for all somatic mutations, not just for hotspots sequenced using the methods reported here. This would also allow further investigation into relationships between driver and passenger mutations through the analysis of different mutation allelic fractions and investigation of subclonal relationships, which may further improve our understanding of the molecular architecture of mCRCs.

Another area of further work that may help to improve the accuracy of the results would be to obtain gene expression data for the COIN and COIN-B cohorts, as this would undoubtedly provide more accurate insights into the eQTL analyses described here due to differences between healthy and diseased tissues in terms of gene expression (Closa et al., 2014).

While GWAS methods have been proven successful in the identification of new genetic variants associated with complex disease phenotypes, they focus on the detection of main effects for each SNP separately (Szymczak et al., 2009). However, according to the CD/CV hypothesis, germline variants associated with common diseases are likely to have modest effects and explain only a small fraction of the overall heritability (Frazer et al., 2009). Therefore, methods that can measure the combined effects of SNPs through analysing them simultaneously may lead to more accurate estimations of the combined effect of SNPs on CRC prognosis.

Advances in technology have seen the emergence of machine learning approaches in the context of genetic association studies, which provide several alternatives for performing multi-SNP analyses, such as penalised regression, decision trees and artificial neural network methods (Szymczak et al., 2009; Kourou et al., 2015; Romagnoni et al., 2019). Machine learning methods have recently been employed in the analysis of several conditions including Crohn's Disease (Romagnoni et al., 2019) and Alzheimer's Disease (Fisher et al., 2019), as well as cancer prognosis (Kourou et al., 2015) and response to therapy (Huang et al., 2018a). It has been shown that machine learning methods can be used to substantially improve the accuracy of predicting cancer susceptibility, recurrence and mortality, and are serving to improve our basic understanding of cancer development and progression (Cruz and Wishart, 2007). An interesting area of further research could be to analyse the COIN and COIN-B genotyping data using a machine learning approach in order to analyse the combined effects of SNPs on prognosis.

7.5 Outlook

The work in this thesis has identified that somatic mutations in the *KRAS*, *BRAF* and *NRAS* oncogenes and tumour MSI status have a significant effect on survival in patients with mCRC. Novel germline variants have also been found to be associated with mCRC prognosis at a level of genome-wide significance. While these SNPs are yet to be successfully validated, it is possible that further analyses may replicate these findings, which could lead to their use as biomarkers for potentially druggable genes.

References

Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., Hurles, M. E., and McVean, G. A. (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467(7319):1061–1073.

Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T., and McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65.

Abuli, A., Lozano, J. J., Rodriguez-Soler, M., Jover, R., Bessa, X., Munoz, J., Esteban-Jurado, C., Fernandez-Rozadilla, C., Carracedo, A., Ruiz-Ponte, C., Cubiella, J., Balaguer, F., Bujanda, L., Rene, J. M., Clofent, J., Morillas, J. D., Nicolas-Perez, D., Xicola, R. M., Llor, X., Pique, J. M., Andreu, M., Castells, A., and Castellvi-Bel, S. (2013). Genetic susceptibility variants associated with colorectal cancer prognosis. *Carcinogenesis*, 34(10):2286–2291.

Adams, R. A., Meade, A. M., Seymour, M. T., Wilson, R. H., Madi, A., Fisher, D., Kenny, S. L., Kay, E., Hodgkinson, E., Pope, M., Rogers, P., Wasan, H., Falk, S., Gollins, S., Hickish, T., Bessell, E. M., Propper, D., Kennedy, M. J., Kaplan, R., and Maughan, T. S. (2011). Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet Oncol*, 12(7):642–653.

Al-Tassan, N., Chmiel, N. H., Maynard, J., Fleming, N., Livingston, A. L., Williams, G. T., Hodges, A. K., Davies, D. R., David, S. S., Sampson, J. R., and Cheadle, J. P. (2002). Inherited variants of MYH associated with somatic G:C–>T:A mutations in colorectal tumors. *Nat Genet*, 30(2):227–232.

Al-Tassan, N. A., Whiffin, N., Hosking, F. J., Palles, C., Farrington, S. M., Dobbins, S. E., Harris, R., Gorman, M., Tenesa, A., Meyer, B. F., Wakil, S. M., Kinnersley, B., Campbell, H., Martin, L., Smith, C. G., Idziaszczyk, S., Barclay, E., Maughan, T. S., Kaplan, R., Kerr, R., Kerr, D., Buchannan, D. D., Ko Win, A., Hopper, J., Jenkins, M., Lindor, N. M., Newcomb, P. A., Gallinger, S., Conti, D., Schumacher, F., Casey, G., Dunlop, M. G., Tomlinson, I. P., Cheadle, J. P., and Houlston, R. S. (2015). A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. *Sci Rep*, 5:10442.

Alhopuro, P., Alazzouzi, H., Sammalkorpi, H., Davalos, V., Salovaara, R., Hemminki, A., Jarvinen, H., Mecklin, J.-P., Schwartz, S. J., Aaltonen, L. A., and Arango, D. (2005). SMAD4 levels and response to 5-fluorouracil in colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 11(17):6311–6316.

Allegra, C. J., Rumble, R. B., Hamilton, S. R., Mangu, P. B., Roach, N., Hantel, A., and Schilsky, R. L. (2016). Extended RAS Gene Mutation Testing in Metastatic Colorectal Carcinoma to Predict Response to Anti-Epidermal Growth Factor Receptor Monoclonal Antibody Therapy: American Society of Clinical Oncology Provisional Clinical Opinion Update 2015. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 34(2):179–185.

Amado, R. G., Wolf, M., Peeters, M., Van Cutsem, E., Siena, S., Freeman, D. J., Juan, T., Sikorski, R., Suggs, S., Radinsky, R., Patterson, S. D., and Chang, D. D. (2008). Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol*, 26(10):1626–1634.

Amaki-Takao, M., Yamaguchi, T., Natsume, S., Iijima, T., Wakaume, R., Takahashi, K., Matsumoto, H., and Miyaki, M. (2016). Colorectal Cancer with BRAF D594G Mutation Is Not Associated with Microsatellite Instability or Poor Prognosis. *Oncology*, 91(3):162–170.

Ambrosini-Spaltro, A., Farnedi, A., Montironi, R., and Foschini, M. P. (2011). IGFBP2 as an immunohistochemical marker for prostatic adenocarcinoma. *Appl Immunohistochem Mol Morphol*, 19(4):318–328.

Amin, M. B., Greene, F. L., Edge, S. B., Compton, C. C., Gershenwald, J. E., Brookland, R. K., Meyer, L., Gress, D. M., Byrd, D. R., and Winchester, D. P. (2017). The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more ”personalized” approach to cancer staging. *CA Cancer J Clin*, 67(2):93–99.

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., and Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nat Protoc*, 5(9):1564–1573.

Andre, T., Boni, C., Mounedji-Boudiaf, L., Navarro, M., Tabernero, J., Hickish, T., Topham, C., Zaninelli, M., Clingan, P., Bridgewater, J., Tabah-Fisch, I., and de Gramont, A. (2004). Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med*, 350(23):2343–2351.

Andreyev, H. J., Norman, A. R., Cunningham, D., Oates, J., Dix, B. R., Iacopetta, B. J., Young, J., Walsh, T., Ward, R., Hawkins, N., Beranek, M., Jandik, P., Benamouzig, R., Jullian, E., Laurent-Puig, P., Olschwang, S., Muller, O., Hoffmann, I., Rabes, H. M., Zietz, C., Troungos, C., Valavanis, C., Yuen, S. T., Ho, J. W., Croke, C. T., O'Donoghue, D. P., Giaretti, W., Rapallo, A., Russo, A., Bazan, V., Tanaka, M., Omura, K., Azuma, T., Ohkusa, T., Fujimori, T., Ono, Y., Pauly, M., Faber, C., Glaesener, R., de Goeij, A. F., Arends, J. W., Andersen, S. N., Lovig, T., Breivik, J., Gaudernack, G., Clausen, O. P., De Angelis, P. D., Meling, G. I., Rognum, T. O., Smith, R., Goh, H. S., Font, A., Rosell, R., Sun, X. F., Zhang, H., Benhattar, J., Losi, L., Lee, J. Q., Wang, S. T., Clarke, P. A., Bell, S., Quirke, P., Bubb, V. J., Piris, J., Cruickshank, N. R., Morton, D., Fox, J. C., Al-Mulla, F., Lees, N., Hall, C. N., Snary, D., Wilkinson, K., Dillon, D., Costa, J., Pricolo, V. E., Finkelstein, S. D., Thebo, J. S., Senagore, A. J., Halter, S. A., Wadler, S., Malik, S., Krtolica, K., and Urosevic, N. (2001). Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study. *Br J Cancer*, 85(5):692–696.

Andreyev, H. J., Norman, A. R., Cunningham, D., Oates, J. R., and Clarke, P. A. (1998). Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *J Natl Cancer Inst*, 90(9):675–684.

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2017). Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, 66(4):683–691.

Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23(10):1294–1296.

Bacolod, M. D. and Barany, F. (2011). Molecular profiling of colon tumors: the search for clinically relevant biomarkers of progression, prognosis, therapeutics, and predisposition. *Ann Surg Oncol*, 18(13):3694–3700.

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B., Ng, P. K., Jeong, K. J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E. M., Mularoni, L., Rubio-Perez, C., Nagarajan, N., Cortes-Ciriano, I., Zhou, D. C., Liang, W. W., Hess, J. M., Yellapantula, V. D., Tamborero, D., Gonzalez-Perez, A., Suphavilai, C., Ko, J. Y., Khurana, E., Park, P. J., Van Allen, E. M., Liang, H., Lawrence, M. S., Godzik, A., Lopez-Bigas, N., Stuart, J., Wheeler, D., Getz, G., Chen, K., Lazar, A. J., Mills, G. B., Karchin, R., and Ding, L. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, 173(2):371–385.e18.

Ball, R. D. (2013). Designing a GWAS: power, sample size, and data structure. *Methods Mol Biol*, 1019:37–98.

Baradaran-Heravi, A., Raams, A., Lubieniecka, J., Cho, K. S., DeHaai, K. A., Basiratnia, M., Mari, P.-O., Xue, Y., Rauth, M., Olney, A. H., Shago, M., Choi, K., Weksberg, R. A., Nowaczyk, M. J. M., Wang, W., Jaspers, N. G. J., and Boerkoel, C. F. (2012). SMARCAL1 deficiency predisposes to non-Hodgkin lymphoma and hypersensitivity to genotoxic agents in vivo. *American journal of medical genetics. Part A*, 158A(9):2204–2213.

Barault, L., Charon-Barra, C., Jooste, V., de la Vega, M. F., Martin, L., Roignot, P., Rat, P., Bouvier, A. M., Laurent-Puig, P., Faivre, J., Chapusot, C., and Piard, F. (2008). Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res*, 68(20):8541–8546.

Battle, A., Brown, C. D., Engelhardt, B. E., and Montgomery, S. B. (2017). Genetic effects on gene expression across human tissues. *Nature*, 550(7675):204–213.

Behjati, S. and Tarpey, P. S. (2013). What is next generation sequencing? *Archives of disease in childhood. Education and practice edition*, 98(6):236–238.

Belanger, C., Speizer, F. E., Hennekens, C. H., Rosner, B., Willett, W., and Bain, C. (1980). The nurses' health study: current findings. *Am J Nurs*, 80(7):1333.

Belanger, C. F., Hennekens, C. H., Rosner, B., and Speizer, F. E. (1978). The nurses' health study. *Am J Nurs*, 78(6):1039–1040.

Benatti, P., Gafa, R., Barana, D., Marino, M., Scarselli, A., Pedroni, M., Maestri, I., Guerzoni, L., Roncucci, L., Menigatti, M., Roncari, B., Maffei, S., Rossi, G., Ponti, G., Santini, A., Losi, L., Di Gregorio, C., Oliani, C., Ponz de Leon, M., and Lanza, G. (2005). Microsatellite instability and colorectal cancer prognosis. *Clin Cancer Res*, 11(23):8332–8340.

Bertagnolli, M. M., Niedzwiecki, D., Compton, C. C., Hahn, H. P., Hall, M., Damas, B., Jewell, S. D., Mayer, R. J., Goldberg, R. M., Saltz, L. B., Warren, R. S., and Redston, M. (2009). Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon cancer: Cancer and Leukemia Group B Protocol 89803. *J Clin Oncol*, 27(11):1814–1821.

Bertotti, A., Papp, E., Jones, S., Adleff, V., Anagnostou, V., Lupo, B., Sausen, M., Phallen, J., Hruban, C. A., Tokheim, C., Niknafs, N., Nesselbush, M., Lytle, K., Sassi, F., Cottino, F., Migliardi, G., Zanella, E. R., Ribero, D., Russolillo, N., Mellano, A., Muratore, A., Paraluppi, G., Salizzoni, M., Marsoni, S., Kragh, M., Lantto, J., Cassingena, A., Li, Q. K., Karchin, R., Scharpf, R., Sartore-Bianchi, A., Siena, S., Diaz Jr., L. A., Trusolino, L., and Velculescu, V. E. (2015). The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature*, 526(7572):263–267.

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Res*, 77(15):3965–3981.

Biau, D. J., Kerneis, S., and Porcher, R. (2008). Statistics in brief: the importance of sample size in the planning and interpretation of medical research. *Clin Orthop Relat Res*, 466(9):2282–2288.

Bokemeyer, C., Bondarenko, I., Hartmann, J. T., de Braud, F., Schuch, G., Zubel, A., Celik, I., Schlichting, M., and Koralewski, P. (2011). Efficacy according to biomarker status of cetuximab plus FOLFOX-4 as first-line treatment for metastatic colorectal cancer: the OPUS study. *Ann Oncol*, 22(7):1535–1546.

Boyle, T., Fritschi, L., Platell, C., and Heyworth, J. (2013). Lifestyle factors associated with survival after colorectal cancer diagnosis. *Br J Cancer*, 109(3):814–822.

Bradburn, M. J., Clark, T. G., Love, S. B., and Altman, D. G. (2003). Survival analysis part II: multivariate data analysis–an introduction to concepts and methods. *Br J Cancer*, 89(3):431–436.

Brenner, H., Kloor, M., and Pox, C. P. (2014). Colorectal cancer. *Lancet*, 383(9927):1490–1502.

Burd, C. E., Liu, W., Huynh, M. V., Waqas, M. A., Gillahan, J. E., Clark, K. S., Fu, K., Martin, B. L., Jeck, W. R., Souroullas, G. P., Darr, D. B., Zedek, D. C., Miley, M. J., Baguley, B. C., Campbell, S. L., and Sharpless, N. E. (2014). Mutation-specific RAS oncogenicity explains NRAS codon 61 selection in melanoma. *Cancer Discov*, 4(12):1418–1429.

Bush, W. S. and Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Comput Biol*, 8(12):e1002822.

Busund, L. T., Richardsen, E., Busund, R., Ukkonen, T., Bjornsen, T., Busch, C., and Stalsberg, H. (2005). Significant expression of IGFBP2 in breast cancer compared with benign lesions. *J Clin Pathol*, 58(4):361–366.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., and Munafo, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*, 14(5):365–376.

Cancer Research UK (2018). UK incidence statistics,.

Cantor, R. M., Lange, K., and Sinsheimer, J. S. (2010). Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am J Hum Genet*, 86(1):6–22.

Cardo-Vila, M., Giordano, R. J., Sidman, R. L., Bronk, L. F., Fan, Z., Mendelsohn, J., Arap, W., and Pasqualini, R. (2010). From combinatorial peptide selection to drug prototype (II): targeting the epidermal growth factor receptor pathway. *Proc Natl Acad Sci U S A*, 107(11):5118–5123.

Carethers, J. M., Koi, M., and Tseng-Rogenski, S. S. (2015). EMAST is a Form of Microsatellite Instability That is Initiated by Inflammation and Modulates Colorectal Cancer Progression. *Genes (Basel)*, 6(2):185–205.

Carithers, L. J. and Moore, H. M. (2015). The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank*, 13(5):307–308.

Carroll, C., Badu-Nkansah, A., Hunley, T., Baradaran-Heravi, A., Cortez, D., and Frangoul, H. (2013). Schimke Immunoosseous Dysplasia associated with undifferentiated carcinoma and a novel SMARCAL1 mutation in a child. *Pediatric blood & cancer*, 60(9):E88–90.

Cassidy, J., Tabernero, J., Twelves, C., Brunet, R., Butts, C., Conroy, T., Debraud, F., Figer, A., Grossmann, J., Sawada, N., Schoffski, P., Sobrero, A., Van Cutsem, E., and Diaz-Rubio, E. (2004). XELOX (capecitabine plus oxaliplatin): active first-line therapy for patients with metastatic colorectal cancer. *J Clin Oncol*, 22(11):2084–2091.

Castro-Giner, F., Ratcliffe, P., and Tomlinson, I. (2015). The mini-driver model of polygenic cancer evolution. *Nature reviews. Cancer*, 15(11):680–685.

Catalano, C., da Silva Filho, M. I., Frank, C., Jiraskova, K., Vymetalkova, V., Levy, M., Liska, V., Vycital, O., Naccarati, A., Vodickova, L., Hemminki, K., Vodicka, P., Weber, A. N. R., and Forsti, A. (2018). Investigation of single and synergic effects of NLRC5 and PD-L1 variants on the risk of colorectal cancer. *PLoS One*, 13(2):e0192385.

Catalano, V., Loupakis, F., Graziano, F., Torresi, U., Bisonni, R., Mari, D., Fornaro, L., Baldelli, A. M., Giordani, P., Rossi, D., Alessandroni, P., Giustini, L., Silva, R. R., Falcone, A., D'Emidio, S., and Fedeli, S. L. (2009). Mucinous histology predicts for poor response rate and overall survival of patients with colorectal cancer and treated with first-line oxaliplatin- and/or irinotecan-based chemotherapy. *British journal of cancer*, 100(6):881–887.

Cayre, A., Rossignol, F., Clottes, E., and Penault-Llorca, F. (2003). aHIF but not HIF-1alpha transcript is a poor prognostic marker in human breast cancer. *Breast Cancer Res*, 5(6):R223–30.

Cercek, A., Braghiroli, M. I., Chou, J. F., Hechtman, J. F., Kemeny, N., Saltz, L., Capanu, M., and Yaeger, R. (2017). Clinical Features and Outcomes of Patients with Colorectal Cancers Harboring NRAS Mutations. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 23(16):4753–4760.

Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., Hirschhorn, J. N., Abecasis, G., Altshuler, D., Bailey-Wilson, J. E., Brooks, L. D., Cardon, L. R., Daly, M., Donnelly, P., Fraumeni Jr., J. F., Freimer, N. B., Gerhard, D. S., Gunter, C., Guttmacher, A. E., Guyer, M. S., Harris, E. L., Hoh, J., Hoover, R., Kong, C. A., Merikangas, K. R., Morton, C. C., Palmer, L. J., Phimister, E. G., Rice, J. P., Roberts, J., Rotimi, C., Tucker, M. A., Vogan, K. J., Wacholder, S., Wijsman, E. M., Winn, D. M., and Collins, F. S. (2007). Replicating genotype-phenotype associations. *Nature*, 447(7145):655–660.

Chen, Y., Yu, X., Xu, Y., and Shen, H. (2017). Identification of dysregulated lncRNAs profiling and metastasis-associated lncRNAs in colorectal cancer by genome-wide analysis. *Cancer Med*, 6(10):2321–2330.

Chionh, F., Lau, D., Yeung, Y., Price, T., and Tebbutt, N. (2017). Oral versus intravenous fluoropyrimidines for colorectal cancer. *Cochrane Database Syst Rev*, 7:Cd008398.

Christodoulidis, G., Spyridakis, M., Symeonidis, D., Kapatou, K., Manolakis, A., and Tepetes, K. (2010). Clinicopathological differences between right- and left-sided colonic tumors and impact upon survival. *Techniques in coloproctology*, 14 Suppl 1:S45–7.

Church, D. N., Stelloo, E., Nout, R. A., Valtcheva, N., Depreeuw, J., ter Haar, N., Noske, A., Amant, F., Tomlinson, I. P., Wild, P. J., Lambrechts, D., Jurgenliemk-Schulz, I. M., Jobsen, J. J., Smit, V. T., Creutzberg, C. L., and Bosse, T. (2015). Prognostic significance of POLE proofreading mutations in endometrial cancer. *J Natl Cancer Inst*, 107(1):402.

Clark, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. *Br J Cancer*, 89(2):232–238.

Closa, A., Cordero, D., Sanz-Pamplona, R., Sole, X., Crous-Bou, M., Pare-Brunet, L., Berenguer, A., Guino, E., Lopez-Doriga, A., Guardiola, J., Biondo, S., Salazar, R., and Moreno, V. (2014). Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis. *Carcinogenesis*, 35(9):2039–2046.

Colditz, G. A., Manson, J. E., and Hankinson, S. E. (1997). The Nurses' Health Study: 20-year contribution to the understanding of health among women. *J Womens Health*, 6(1):49–62.

Couch, F. B., Bansbach, C. E., Driscoll, R., Luzwick, J. W., Glick, G. G., Betous, R., Carroll, C. M., Jung, S. Y., Qin, J., Cimprich, K. A., and Cortez, D. (2013). ATR phosphorylates SMARCAL1 to prevent replication fork collapse. *Genes Dev*, 27(14):1610–1623.

Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J., and Der, C. J. (2014). Drugging the undruggable RAS: Mission possible? *Nature reviews. Drug discovery*, 13(11):828–851.

Cox, K. E., Marechal, A., and Flynn, R. L. (2016). SMARCAL1 Resolves Replication Stress at ALT Telomeres. *Cell Rep*, 14(5):1032–1040.

Coyle, C., Cafferty, F. H., Rowley, S., MacKenzie, M., Berkman, L., Gupta, S., Pramesh, C. S., Gilbert, D., Kynaston, H., Cameron, D., Wilson, R. H., Ring, A., and Langley, R. E. (2016). ADD-ASPIRIN: A phase III, double-blind, placebo controlled, randomised trial assessing the effects of aspirin on disease recurrence and survival after primary therapy in common non-metastatic solid tumours. *Contemp Clin Trials*, 51:56–64.

Cremolini, C., Di Bartolomeo, M., Amatu, A., Antoniotti, C., Moretto, R., Berenato, R., Perrone, F., Tamborini, E., Aprile, G., Lonardi, S., Sartore-Bianchi, A., Fontanini, G., Milione, M., Lauricella, C., Siena, S., Falcone, A., de Braud, F., Loupakis, F., and Pietrantonio, F. (2015). BRAF codons 594 and 596 mutations identify a new molecular subtype of metastatic colorectal cancer at favorable prognosis. *Ann Oncol*.

Cruz, J. A. and Wishart, D. S. (2007). Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, 2:59–77.

Dahlin, A. M., Palmqvist, R., Henriksson, M. L., Jacobsson, M., Eklof, V., Rutegard, J., Oberg, A., and Van Guelpen, B. R. (2010). The role of the CpG island methylator phenotype in colorectal cancer prognosis depends on microsatellite instability screening status. *Clin Cancer Res*, 16(6):1845–1855.

Dai, J., Gu, J., Huang, M., Eng, C., Kopetz, E. S., Ellis, L. M., Hawk, E., and Wu, X. (2012). GWAS-identified colorectal cancer susceptibility loci associated with clinical outcomes. *Carcinogenesis*, 33(7):1327–1331.

de Bakker, P. I., Ferreira, M. A., Jia, X., Neale, B. M., Raychaudhuri, S., and Voight, B. F. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*, 17(R2):R122–8.

de Gramont, A., Figer, A., Seymour, M., Homerin, M., Hmissi, A., Cassidy, J., Boni, C., Cortes-Funes, H., Cervantes, A., Freyer, G., Papamichael, D., Le Bail, N., Louvet, C., Hendler, D., de Braud, F., Wilson, C., Morvan, F., and Bonetti, A. (2000). Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J Clin Oncol*, 18(16):2938–2947.

De Roock, W., Claes, B., Bernasconi, D., De Schutter, J., Biesmans, B., Fountzilas, G., Kalogeras, K. T., Kotoula, V., Papamichael, D., Laurent-Puig, P., Penault-Llorca, F., Rougier, P., Vincenzi, B., Santini, D., Tonini, G., Cappuzzo, F., Frattini, M., Molinari, F., Saletti, P., De Dosso, S., Martini, M., Bardelli, A., Siena, S., Sartore-Bianchi, A., Tabernero, J., Macarulla, T., Di Fiore, F., Gangloff, A. O., Ciardiello, F., Pfeiffer, P., Qvortrup, C., Hansen, T. P., Van Cutsem, E., Piessevaux, H., Lambrechts, D., Delorenzi, M., and Tejpar, S. (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):753–762.

De Roock, W., Jonker, D. J., Di Nicolantonio, F., Sartore-Bianchi, A., Tu, D., Siena, S., Lamba, S., Arena, S., Frattini, M., Piessevaux, H., Van Cutsem, E., O'Callaghan, C. J., Khambata-Ford, S., Zalcberg, J. R., Simes, J., Karapetis, C. S., Bardelli, A., and Tejpar, S. (2010b). Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab. *JAMA*, 304(16):1812–1820.

De Sousa, E. M. F., Wang, X., Jansen, M., Fessler, E., Trinh, A., de Rooij, L. P., de Jong, J. H., de Boer, O. J., van Leersum, R., Bijlsma, M. F., Rodermond, H., van der Heijden, M., van Noesel, C. J., Tuynman, J. B., Dekker, E., Markowetz, F., Medema, J. P., and Vermeulen, L. (2013). Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med*, 19(5):614–618.

Demunter, A., Stas, M., Degreef, H., De Wolf-Peeters, C., and van den Oord, J. J. (2001). Analysis of N- and K-ras mutations in the distinctive tumor progression phases of melanoma. *The Journal of investigative dermatology*, 117(6):1483–1489.

Domingo, E., Freeman-Mills, L., Rayner, E., Glaire, M., Briggs, S., Vermeulen, L., Fessler, E., Medema, J. P., Boot, A., Morreau, H., van Wezel, T., Liefers, G. J., Lothe, R. A., Danielsen, S. A., Sveen, A., Nesbakken, A., Zlobec, I., Lugli, A., Koelzer, V. H., Berger, M. D., Castellvi-Bel, S., Munoz, J., de Bruyn, M., Nijman, H. W., Novelli, M., Lawson, K., Oukrif, D., Frangou, E., Dutton, P., Tejpar, S., Delorenzi, M., Kerr, R., Kerr, D., Tomlinson, I., and Church, D. N. (2016). Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *Lancet Gastroenterol Hepatol*, 1(3):207–216.

Dotor, E., Cuatrecases, M., Martinez-Iniesta, M., Navarro, M., Vilardell, F., Guino, E., Pareja, L., Figueras, A., Mollevi, D. G., Serrano, T., de Oca, J., Peinado, M. A., Moreno, V., Germa, J. R., Capella, G., and Villanueva, A. (2006). Tumor thymidylate synthase 1494del6 genotype as a prognostic factor in colorectal cancer patients receiving fluorouracil-based adjuvant treatment. *J Clin Oncol*, 24(10):1603–1611.

Douillard, J. Y., Cunningham, D., Roth, A. D., Navarro, M., James, R. D., Karasek, P., Jandik, P., Iveson, T., Carmichael, J., Alakl, M., Gruia, G., Awad, L., and Rougier, P. (2000). Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet*, 355(9209):1041–1047.

Douillard, J. Y., Oliner, K. S., Siena, S., Tabernero, J., Burkes, R., Barugel, M., Humblet, Y., Bodoky, G., Cunningham, D., Jassem, J., Rivera, F., Kocakova, I., Ruff, P., Blasinska-Morawiec, M., Smakal, M., Canon, J. L., Rother, M., Williams, R., Rong, A., Wiezorek, J., Sidhu, R., and Patterson, S. D. (2013). Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med*, 369(11):1023–1034.

Douillard, J. Y., Siena, S., Cassidy, J., Tabernero, J., Burkes, R., Barugel, M., Humblet, Y., Bodoky, G., Cunningham, D., Jassem, J., Rivera, F., Kocakova, I., Ruff, P., Blasinska-Morawiec, M., Smakal, M., Canon, J. L., Rother, M., Oliner, K. S., Wolf, M., and Gansert, J. (2010). Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *J Clin Oncol*, 28(31):4697–4705.

Draht, M. X. G., Goudkade, D., Koch, A., Grabsch, H. I., Weijenberg, M. P., van Engeland, M., Melotte, V., and Smits, K. M. (2018). Prognostic DNA methylation markers for sporadic colorectal cancer: a systematic review. *Clin Epigenetics*, 10:35.

Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G., Struewing, J. P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C. S., Bowman, R., Meyer, K. B., Haiman, C. A., Kolonel, L. K., Henderson, B. E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C. Y., Wu, P. E., Wang, H. C., Eccles, D., Evans, D. G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M. R., Rahman, N., Chenevix-Trench, G., Bojesen, S. E., Nordestgaard, B. G., Axelsson, C. K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Hunter, D. J., Hankinson, S. E., Cox, D. G., Hall, P., Wedren, S., Liu, J., Low, Y. L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R. A., Jacobi, C. E., Devilee, P., Klijn, J. G., Sigurdson, A. J., Doody, M. M., Alexander, B. H., Zhang, J., Cox, A., Brock, I. W., MacPherson, G., Reed, M. W., Couch, F. J., Goode, E. L., Olson, J. E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R. L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J. L., McCredie, M., Southey, M., Giles, G. G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y. D., Spurdle, A. B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V. M., Kataja, V., Hartikainen, J., Day, N. E., Cox, D. R., and Ponder, B. A. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447(7148):1087–1093.

Eklof, V., Wikberg, M. L., Edin, S., Dahlin, A. M., Jonsson, B. A., Oberg, A., Rutegard, J., and Palmqvist, R. (2013). The prognostic role of KRAS, BRAF, PIK3CA and PTEN in colorectal cancer. *Br J Cancer*, 108(10):2153–2163.

el Atiq, F., Garrouste, F., Remacle-Bonnet, M., Sastre, B., and Pommier, G. (1994). Alterations in serum levels of insulin-like growth factors and insulin-like growth-factor-binding proteins in patients with colorectal cancer. *Int J Cancer*, 57(4):491–497.

Engelmann, B. E., Høgdall, E., Holländer, N. H., and Iveson, T. (2016). TransSCOT – translational research based on the Short Course Oncology Therapy (SCOT) cohort. *Annals of Oncology*, 27(suppl_6).

Erichsen, H. C. and Chanock, S. J. (2004). SNPs in cancer research and treatment. *Br J Cancer*, 90(4):747–751.

Esteller, M. (2008). Epigenetics in cancer. *N Engl J Med*, 358(11):1148–1159.

Fadista, J., Manning, A. K., Florez, J. C., and Groop, L. (2016). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet*, 24(8):1202–1205.

Fan, R., Huang, C. H., Lo, S. H., Zheng, T., and Ionita-Laza, I. (2011). Identifying rare disease variants in the Genetic Analysis Workshop 17 simulated data: a comparison of several statistical approaches. *BMC Proc*, 5 Suppl 9:S17.

Farina-Sarasqueta, A., van Lijnschoten, G., Moerland, E., Creemers, G. J., Lemmens, V. E., Rutten, H. J., and van den Brule, A. J. (2010). The BRAF V600E mutation is an independent prognostic factor for survival in stage II and stage III colon cancer patients. *Ann Oncol*, 21(12):2396–2402.

Faron, M., Pignon, J. P., Malka, D., Bourredjem, A., Douillard, J. Y., Adenis, A., Elias, D., Bouche, O., and Ducreux, M. (2015). Is primary tumour resection associated with survival improvement in patients with colorectal cancer and unresectable synchronous metastases? A pooled analysis of individual data from four randomised trials. *Eur J Cancer*, 51(2):166–176.

Fearnhead, N. S., Britton, M. P., and Bodmer, W. F. (2001). The ABC of APC. *Hum Mol Genet*, 10(7):721–733.

Fearon, E. R. (2011). Molecular genetics of colorectal cancer. *Annu Rev Pathol*, 6:479–507.

Fearon, E. R. and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, 61(5):759–767.

Fisher, C. K., Smith, A. M., and Walsh, J. R. (2019). Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Scientific reports*, 9(1):13622.

Foxtrot Collaborative Group (2012). Feasibility of preoperative chemotherapy for locally advanced, operable colon cancer: the pilot phase of a randomised controlled trial. *Lancet Oncol*, 13(11):1152–1160.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., de Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Altshuler, D., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Tsunoda, T., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Gibbs, R. A., Belmont, J. W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Wheeler, D. A., Yakub, I., Gabriel, S. B., Onofrio, R. C., Richter, D. J., Ziaugra, L., Birren, B. W., Daly, M. J., Altshuler, D., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archeveque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R., and Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164):851–861.

Frazer, K. A., Murray, S. S., Schork, N. J., and Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nature reviews. Genetics*, 10(4):241–251.

Fukushima, T. and Kataoka, H. (2007). Roles of insulin-like growth factor binding protein-2 (IGFBP-2) in glioblastoma. *Anticancer Res*, 27(6a):3685–3692.

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pages, C., Tosolini, M., Camus, M., Berger, A., Wind, P., Zinzindohoue, F., Bruneval, P., Cugnenc, P. H., Trajanoski, Z., Fridman, W. H., and Pages, F. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*, 313(5795):1960–1964.

Gao, S., Sun, Y., Zhang, X., Hu, L., Liu, Y., Chua, C. Y., Phillips, L. M., Ren, H., Fleming, J. B., Wang, H., Chiao, P. J., Hao, J., and Zhang, W. (2016). IGFBP2 Activates the NF-kappaB Pathway to Drive Epithelial-Mesenchymal Transition and Invasive Character in Pancreatic Ductal Adenocarcinoma. *Cancer Res*, 76(22):6543–6554.

Garcia-Albeniz, X., Nan, H., Valeri, L., Morikawa, T., Kuchiba, A., Phipps, A. I., Hutter, C. M., Peters, U., Newcomb, P. A., Fuchs, C. S., Giovannucci, E. L., Ogino, S., and Chan, A. T. (2013). Phenotypic and tumor molecular characterization of colorectal cancer in relation to a susceptibility SMAD7 variant associated with survival. *Carcinogenesis*, 34(2):292–298.

Garner, C. (2007). Upward bias in odds ratio estimates from genome-wide association studies. *Genetic epidemiology*, 31(4):288–295.

Goldstein, N. S. and Armin, M. (2001). Epidermal growth factor receptor immunohistochemical reactivity in patients with American Joint Committee on Cancer Stage IV colon adenocarcinoma: implications for a standardized scoring system. *Cancer*, 92(5):1331–1346.

Gonsalves, W. I., Mahoney, M. R., Sargent, D. J., Nelson, G. D., Alberts, S. R., Sinicrope, F. A., Goldberg, R. M., Limburg, P. J., Thibodeau, S. N., Grothey, A., Hubbard, J. M., Chan, E., Nair, S., Berenberg, J. L., and McWilliams, R. R. (2014). Patient and tumor characteristics and BRAF and KRAS mutations in colon cancer, NCCTG/Alliance N0147. *J Natl Cancer Inst*, 106(7).

Gray, R., Barnwell, J., McConkey, C., Hills, R. K., Williams, N. S., and Kerr, D. J. (2007). Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet*, 370(9604):2020–2029.

Griffith, O. L., Montgomery, S. B., Bernier, B., Chu, B., Kasaian, K., Aerts, S., Mahony, S., Sleumer, M. C., Bilenky, M., Haeussler, M., Griffith, M., Gallo, S. M., Giardine, B., Hooghe, B., Van Loo, P., Blanco, E., Ticoll, A., Lithwick, S., Portales-Casamar, E., Donaldson, I. J., Robertson, G., Wadelius, C., De Bleser, P., Vlieghe, D., Halfon, M. S., Wasserman, W., Hardison, R., Bergman, C. M., and Jones, S. J. (2008). ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res*, 36(Database issue):D107–13.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M., and al., E. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. *Cell*, 66(3):589–600.

Gunnlaugsson, A., Anderson, H., Fernebro, E., Kjellen, E., Bystrom, P., Berglund, K., Ekelund, M., Pahlman, L., Holm, T., Glimelius, B., and Johnsson, A. (2009). Multicentre phase II trial of capecitabine and oxaliplatin in combination with radiotherapy for unresectable colorectal cancer: the CORGI-L Study. *Eur J Cancer*, 45(5):807–813.

Gutschner, T. and Diederichs, S. (2012). The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol*, 9(6):703–719.

Gutschner, T., Richtig, G., Haemmerle, M., and Pichler, M. (2018). From biomarkers to therapeutic targets-the promises and perils of long non-coding RNAs in cancer. *Cancer Metastasis Rev*, 37(1):83–105.

Haigis, K. M., Kendall, K. R., Wang, Y., Cheung, A., Haigis, M. C., Glickman, J. N., Niwa-Kawakita, M., Sweet-Cordero, A., Sebolt-Leopold, J., Shannon, K. M., Settleman, J., Giovannini, M., and Jacks, T. (2008). Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. *Nature genetics*, 40(5):600–608.

Haiman, C. A., Chen, G. K., Vachon, C. M., Canzian, F., Dunning, A., Millikan, R. C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C. B., Baglietto, L., Balleine, R., Bandera, E. V., Beckmann, M. W., Berg, C. D., Bernstein, L., Blomqvist, C., Blot, W. J., Brauch, H., Buring, J. E., Carey, L. A., Carpenter, J. E., Chang-Claude, J., Chanock, S. J., Chasman, D. I., Clarke, C. L., Cox, A., Cross, S. S., Deming, S. L., Diasio, R. B., Dimopoulos, A. M., Driver, W. R., Dunnebier, T., Durcan, L., Eccles, D., Edlund, C. K., Ekici, A. B., Fasching, P. A., Feigelson, H. S., Flesch-Janys, D., Fostira, F., Forsti, A., Fountzilas, G., Gerty, S. M., Giles, G. G., Godwin, A. K., Goodfellow, P., Graham, N., Greco, D., Hamann, U., Hankinson, S. E., Hartmann, A., Hein, R., Heinz, J., Holbrook, A., Hoover, R. N., Hu, J. J., Hunter, D. J., Ingles, S. A., Irwanto, A., Ivanovich, J., John, E. M., Johnson, N., Jukkola-Vuorinen, A., Kaaks, R., Ko, Y. D., Kolonel, L. N., Konstantopoulou, I., Kosma, V. M., Kulkarni, S., Lambrechts, D., Lee, A. M., Marchand, L. L., Lesnick, T., Liu, J., Lindstrom, S., Mannermaa, A., Margolin, S., Martin, N. G., Miron, P., Montgomery, G. W., Nevanlinna, H., Nickels, S., Nyante, S., Olswold, C., Palmer, J., Pathak, H., Pectasides, D., Perou, C. M., Peto, J., Pharoah, P. D., Pooler, L. C., Press, M. F., Pylkas, K., Rebbeck, T. R., Rodriguez-Gil, J. L., Rosenberg, L., Ross, E., Rudiger, T., Silva Idos, S., Sawyer, E., Schmidt, M. K., Schulz-Wendtland, R., Schumacher, F., Severi, G., Sheng, X., Signorello, L. B., Sinn, H. P., Stevens, K. N., Southey, M. C., Tapper, W. J., Tomlinson, I., Hogervorst, F. B., Wauters, E., Weaver, J., Wildiers, H., Winqvist, R., Van Den Berg, D., Wan, P., Xia, L. Y., Yannoukakos, D., Zheng, W., Ziegler, R. G., Siddiq, A., Slager, S. L., Stram, D. O., Easton, D., Kraft, P., Henderson, B. E., and Couch, F. J. (2011). A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet*, 43(12):1210–1214.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674.

Hawk, E. T. and Levin, B. (2005). Colorectal cancer prevention. *J Clin Oncol*, 23(2):378–391.

Haydon, A. M. and Jass, J. R. (2002). Emerging pathways in colorectal-cancer development. *Lancet Oncol*, 3(2):83–88.

Haydon, A. M., Macinnis, R. J., English, D. R., and Giles, G. G. (2006). Effect of physical activity and body size on survival after diagnosis with colorectal cancer. *Gut*, 55(1):62–67.

He, Y., Meng, X. M., Huang, C., Wu, B. M., Zhang, L., Lv, X. W., and Li, J. (2014). Long noncoding RNAs: Novel insights into hepatocelluar carcinoma. *Cancer Lett*, 344(1):20–27.

Hennekens, C. H., Buring, J. E., Manson, J. E., Stampfer, M., Rosner, B., Cook, N. R., Belanger, C., LaMotte, F., Gaziano, J. M., Ridker, P. M., Willett, W., and Peto, R. (1996). Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *The New England journal of medicine*, 334(18):1145–1149.

Henriksen, L., Grandal, M. V., Knudsen, S. L., van Deurs, B., and Grovdal, L. M. (2013). Internalization mechanisms of the epidermal growth factor receptor after activation with different ligands. *PLoS One*, 8(3):e58148.

Hinrichs, A. L., Larkin, E. K., and Suarez, B. K. (2009). Population stratification and patterns of linkage disequilibrium. *Genet Epidemiol*, 33 Suppl 1:S88–92.

Hirschhorn, J. N. and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 6(2):95–108.

Hobbs, G. A. and Der, C. J. (2019). RAS Mutations Are Not Created Equal. *Cancer discovery*, 9(6):696–698.

Hobbs, G. A., Der, C. J., and Rossman, K. L. (2016a). RAS isoforms and mutations in cancer at a glance. *Journal of cell science*, 129(7):1287–1292.

Hobbs, G. A., Wittinghofer, A., and Der, C. J. (2016b). Selective Targeting of the KRAS G12C Mutant: Kicking KRAS When It's Down. *Cancer cell*, 29(3):251–253.

Hoggart, C. J., Clark, T. G., De Iorio, M., Whittaker, J. C., and Balding, D. J. (2008). Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol*, 32(2):179–185.

Hoskins, J. M., Ong, P. S., Keku, T. O., Galanko, J. A., Martin, C. F., Coleman, C. A., Wolfe, M., Sandler, R. S., and McLeod, H. L. (2012). Association of eleven common, low-penetrance colorectal cancer susceptibility genetic variants at six risk loci with clinical outcome. *PLoS One*, 7(7):e41954.

Houlston, R. S., Cheadle, J., Dobbins, S. E., Tenesa, A., Jones, A. M., Howarth, K., Spain, S. L., Broderick, P., Domingo, E., Farrington, S., Prendergast, J. G., Pittman, A. M., Theodoratou, E., Smith, C. G., Olver, B., Walther, A., Barnetson, R. A., Churchman, M., Jaeger, E. E., Penegar, S., Barclay, E., Martin, L., Gorman, M., Mager, R., Johnstone, E., Midgley, R., Niittymaki, I., Tuupanen, S., Colley, J., Idziaszczyk, S., Thomas, H. J., Lucassen, A. M., Evans, D. G., Maher, E. R., Maughan, T., Dimas, A., Dermitzakis, E., Cazier, J. B., Aaltonen, L. A., Pharoah, P., Kerr, D. J., Carvajal-Carmona, L. G., Campbell, H., Dunlop, M. G., and Tomlinson, I. P. (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):973–977.

Houlston, R. S., Webb, E., Broderick, P., Pittman, A. M., Di Bernardo, M. C., Lubbe, S., Chandler, I., Vijayakrishnan, J., Sullivan, K., Penegar, S., Carvajal-Carmona, L., Howarth, K., Jaeger, E., Spain, S. L., Walther, A., Barclay, E., Martin, L., Gorman, M., Domingo, E., Teixeira, A. S., Kerr, D., Cazier, J. B., Niittymaki, I., Tuupanen, S., Karhu, A., Aaltonen, L. A., Tomlinson, I. P., Farrington, S. M., Tenesa, A., Prendergast, J. G., Barnetson, R. A., Cetnarskyj, R., Porteous, M. E., Pharoah, P. D., Koessler, T., Hampe, J., Buch, S., Schafmayer, C., Tepel, J., Schreiber, S., Volzke, H., Chang-Claude, J., Hoffmeister, M., Brenner, H., Zanke, B. W., Montpetit, A., Hudson, T. J., Gallinger, S., Campbell, H., and Dunlop, M. G. (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet*, 40(12):1426–1435.

Howie, B. N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*, 5(6):e1000529.

Hu, Z., Wu, C., Shi, Y., Guo, H., Zhao, X., Yin, Z., Yang, L., Dai, J., Hu, L., Tan, W., Li, Z., Deng, Q., Wang, J., Wu, W., Jin, G., Jiang, Y., Yu, D., Zhou, G., Chen, H., Guan, P., Chen, Y., Shu, Y., Xu, L., Liu, X., Liu, L., Xu, P., Han, B., Bai, C., Zhao, Y., Zhang, H., Yan, Y., Ma, H., Chen, J., Chu, M., Lu, F., Zhang, Z., Chen, F., Wang, X., Jin, L., Lu, J., Zhou, B., Lu, D., Wu, T., Lin, D., and Shen, H. (2011). A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genet*, 43(8):792–796.

Huang, C., Clayton, E. A., Matyunina, L. V., McDonald, L. D., Benigno, B. B., Vannberg, F., and McDonald, J. F. (2018a). Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. *Scientific reports*, 8(1):16444.

Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J. L., Danecek, P., Malerba, G., Trabetti, E., Zheng, H. F., Gambaro, G., Richards, J. B., Durbin, R., Timpson, N. J., Marchini, J., and Soranzo, N. (2015). Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun*, 6:8111.

Huang, Q. Q., Ritchie, S. C., Brozynska, M., and Inouye, M. (2018b). Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic Acids Res*, 46(22):e133.

Huels, D. J., Bruens, L., Hodder, M. C., Cammareri, P., Campbell, A. D., Ridgway, R. A., Gay, D. M., Solar-Abboud, M., Faller, W. J., Nixon, C., Zeiger, L. B., McLaughlin, M. E., Morrissey, E., Winton, D. J., Snippert, H. J., van Rheenen, J., and Sansom, O. J. (2018). Wnt ligands influence tumour initiation by controlling the number of intestinal stem cells. *Nat Commun*, 9(1):1132.

Hugen, N., van de Velde, C. J. H., de Wilt, J. H. W., and Nagtegaal, I. D. (2014). Metastatic pattern in colorectal cancer is strongly influenced by histological subtype. *Annals of oncology : official journal of the European Society for Medical Oncology*, 25(3):651–657.

Huiskens, J., van Gulik, T. M., van Lienden, K. P., Engelbrecht, M. R., Meijer, G. A., van Grieken, N. C., Schriek, J., Keijser, A., Mol, L., Molenaar, I. Q., Verhoef, C., de Jong, K. P., Dejong, K. H., Kazemier, G., Ruers, T. M., de Wilt, J. H., van Tinteren, H., and Punt, C. J. (2015). Treatment strategies in colorectal cancer patients with initially unresectable liver-only metastases, a study protocol of the randomised phase 3 CAIRO5 study of the Dutch Colorectal Cancer Group (DCCG). *BMC Cancer*, 15:365.

Hulur, I., Gamazon, E. R., Skol, A. D., Xicola, R. M., Llor, X., Onel, K., Ellis, N. A., and Kupfer, S. S. (2015). Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. *BMC Genomics*, 16:138.

Hutchins, G., Southward, K., Handley, K., Magill, L., Beaumont, C., Stahlschmidt, J., Richman, S., Chambers, P., Seymour, M., Kerr, D., Gray, R., and Quirke, P. (2011). Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. *J Clin Oncol*, 29(10):1261–1270.

Huxley, R. R., Ansary-Moghaddam, A., Clifton, P., Czernichow, S., Parr, C. L., and Woodward, M. (2009). The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. *Int J Cancer*, 125(1):171–180.

Hyman, D. M., Puzanov, I., Subbiah, V., Faris, J. E., Chau, I., Blay, J.-Y., Wolf, J., Raje, N. S., Diamond, E. L., Hollebecque, A., Gervais, R., Elez-Fernandez, M. E., Italiano, A., Hofheinz, R.-D., Hidalgo, M., Chan, E., Schuler, M., Lasserre, S. F., Makrutzki, M., Sirzen, F., Veronese, M. L., Tabernero, J., and Baselga, J. (2015). Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. *The New England journal of medicine*, 373(8):726–736.

Ikenoue, T., Hikiba, Y., Kanai, F., Tanaka, Y., Imamura, J., Imamura, T., Ohta, M., Ijichi, H., Tateishi, K., Kawakami, T., Aragaki, J., Matsumura, M., Kawabe, T., and Omata, M. (2003). Functional analysis of mutations within the kinase activation segment of B-Raf in human colorectal tumors. *Cancer Res*, 63(23):8132–8137.

Imamura, Y., Lochhead, P., Yamauchi, M., Kuchiba, A., Qian, Z. R., Liao, X., Nishihara, R., Jung, S., Wu, K., Nosho, K., Wang, Y. E., Peng, S., Bass, A. J., Haigis, K. M., Meyerhardt, J. A., Chan, A. T., Fuchs, C. S., and Ogino, S. (2014). Analyses of clinicopathological, molecular, and prognostic associations of KRAS codon 61 and codon 146 mutations in colorectal cancer: cohort study and literature review. *Mol Cancer*, 13:135.

Imamura, Y., Morikawa, T., Liao, X., Lochhead, P., Kuchiba, A., Yamauchi, M., Qian, Z. R., Nishihara, R., Meyerhardt, J. A., Haigis, K. M., Fuchs, C. S., and Ogino, S. (2012). Specific mutations in KRAS codons 12 and 13, and patient prognosis in 1075 BRAF wild-type colorectal cancers. *Clin Cancer Res*, 18(17):4753–4763.

Irahara, N., Baba, Y., Nosho, K., Shima, K., Yan, L., Dias-Santagata, D., Iafrate, A. J., Fuchs, C. S., Haigis, K. M., and Ogino, S. (2010). NRAS mutations are rare in colorectal cancer. *Diagnostic molecular pathology : the American journal of surgical pathology, part B*, 19(3):157–163.

Iveson, T. J., Kerr, R. S., Saunders, M. P., Cassidy, J., Hollander, N. H., Tabernero, J., Haydon, A., Glimelius, B., Harkin, A., Allan, K., McQueen, J., Scudder, C., Boyd, K. A., Briggs, A., Waterston, A., Medley, L., Wilson, C., Ellis, R., Essapen, S., Dhadda, A. S., Harrison, M., Falk, S., Raouf, S., Rees, C., Olesen, R. K., Propper, D., Bridgewater, J., Azzabi, A., Farrugia, D., Webb, A., Cunningham, D., Hickish, T., Weaver, A., Gollins, S., Wasan, H. S., and Paul, J. (2018). 3 versus 6 months of adjuvant oxaliplatin-fluoropyrimidine combination therapy for colorectal cancer (SCOT): an international, randomised, phase 3, non-inferiority trial. *Lancet Oncol*, 19(4):562–578.

Iwai, Y., Hamanishi, J., Chamoto, K., and Honjo, T. (2017). Cancer immunotherapies targeting the PD-1 signaling pathway. *Journal of biomedical science*, 24(1):26.

Jannot, A. S., Ehret, G., and Perneger, T. (2015). P $<$ 5 x 10(-8) has emerged as a standard of statistical significance for genome-wide association studies. *J Clin Epidemiol*, 68(4):460–465.

Jasperson, K. W., Tuohy, T. M., Neklason, D. W., and Burt, R. W. (2010). Hereditary and familial colon cancer. *Gastroenterology*, 138(6):2044–2058.

Jass, J. R. (2007). Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*, 50(1):113–130.

Jess, T., Rungoe, C., and Peyrin-Biroulet, L. (2012). Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. *Clin Gastroenterol Hepatol*, 10(6):639–645.

Jin, G., Zheng, S. L., Lilja, H., Kim, S. T., Tao, S., Gao, Z., Young, T., Wiklund, F., Feng, J., Isaacs, W. B., Rittmaster, R. S., Gronberg, H., Condreay, L. D., Sun, J., and Xu, J. (2013). Genome-wide association study identifies loci at ATF7IP and KLK2 associated with percentage of circulating free PSA. *Neoplasia*, 15(1):95–101.

Jones, J. C., Renfro, L. A., Al-Shamsi, H. O., Schrock, A. B., Rankin, A., Zhang, B. Y., Kasi, P. M., Voss, J. S., Leal, A. D., Sun, J., Ross, J., Ali, S. M., Hubbard, J. M., Kipp, B. R., McWilliams, R. R., Kopetz, S., Wolff, R. A., and Grothey, A. (2017). (Non-V600) BRAF Mutations Define a Clinically Distinct Molecular Subtype of Metastatic Colorectal Cancer. *J Clin Oncol*, 35(23):2624–2630.

Juo, Y. Y., Johnston, F. M., Zhang, D. Y., Juo, H. H., Wang, H., Pappou, E. P., Yu, T., Easwaran, H., Baylin, S., van Engeland, M., and Ahuja, N. (2014). Prognostic value of CpG island methylator phenotype among colorectal cancer patients: a systematic review and meta-analysis. *Ann Oncol*, 25(12):2314–2327.

Kang, H. P., Morgan, A. A., Chen, R., Schadt, E. E., and Butte, A. J. (2012). Coanalysis of GWAS with eQTLs reveals disease-tissue associations. *AMIA Jt Summits Transl Sci Proc*, 2012:35–41.

Kaplan, R., Maughan, T., Crook, A., Fisher, D., Wilson, R., Brown, L., and Parmar, M. (2013). Evaluating many treatments and biomarkers in oncology: a new design. *J Clin Oncol*, 31(36):4562–4568.

Karapetis, C. S., Khambata-Ford, S., Jonker, D. J., O'Callaghan, C. J., Tu, D., Tebbutt, N. C., Simes, R. J., Chalchal, H., Shapiro, J. D., Robitaille, S., Price, T. J., Shepherd, L., Au, H. J., Langer, C., Moore, M. J., and Zalcberg, J. R. (2008). K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med*, 359(17):1757–1765.

Karssen, L. C., van Duijn, C. M., and Aulchenko, Y. S. (2016). The GenABEL Project for statistical genomics. *F1000Res*, 5:914.

Kattan, M. W., Hess, K. R., Amin, M. B., Lu, Y., Moons, K. G., Gershenwald, J. E., Gimotty, P. A., Guinney, J. H., Halabi, S., Lazar, A. J., Mahar, A. L., Patel, T., Sargent, D. J., Weiser, M. R., and Compton, C. (2016). American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin*, 66(5):370–374.

Kemeny, N. and Braun Jr., D. W. (1983). Prognostic factors in advanced colorectal carcinoma. Importance of lactic dehydrogenase level, performance status, and white blood cell count. *Am J Med*, 74(5):786–794.

Kerr, R. S., Love, S., Segelov, E., Johnstone, E., Falcon, B., Hewett, P., Weaver, A., Church, D., Scudder, C., Pearson, S., Julier, P., Pezzella, F., Tomlinson, I., Domingo, E., and Kerr, D. J. (2016). Adjuvant capecitabine plus bevacizumab versus capecitabine alone in patients with colorectal cancer (QUASAR 2): an open-label, randomised phase 3 trial. *The Lancet. Oncology*, 17(11):1543–1557.

Kim, J. G., Chae, Y. S., Sohn, S. K., Cho, Y. Y., Moon, J. H., Park, J. Y., Jeon, S. W., Lee, I. T., Choi, G. S., and Jun, S. H. (2008). Vascular endothelial growth factor gene polymorphisms associated with prognosis for patients with colorectal cancer. *Clin Cancer Res*, 14(1):62–66.

Kim, S. H., Park, K. H., Shin, S. J., Lee, K. Y., Kim, T. I., Kim, N. K., Rha, S. Y., and Ahn, J. B. (2018). CpG Island Methylator Phenotype and Methylation of Wnt Pathway Genes Together Predict Survival in Patients with Colorectal Cancer. *Yonsei Med J*, 59(5):588–594.

Kim, S. K., Jung, W. H., and Koo, J. S. (2014). Differential expression of enzymes associated with serine/glycine metabolism in different breast cancer subtypes. *PLoS One*, 9(6):e101004.

Kinzler, K. W. and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell*, 87(2):159–170.

Kinzler, K. W. and Vogelstein, B. (1997). Cancer-susceptibility genes. Gatekeepers and caretakers.

Kirk, R. (2011). Genetics: In colorectal cancer, not all KRAS mutations are created equal. *Nature reviews. Clinical oncology*, 8(1):1.

Klein, R. J. (2007). Power analysis for genome-wide association studies. *BMC Genet*, 8:58.

Knudson, A. G. (1996). Hereditary cancer: two hits revisited. *J Cancer Res Clin Oncol*, 122(3):135–140.

Knudson, A. G. (1997). Hereditary predisposition to cancer. *Ann N Y Acad Sci*, 833:58–67.

Kohne, C. H., Cunningham, D., Di Costanzo, F., Glimelius, B., Blijham, G., Aranda, E., Scheithauer, W., Rougier, P., Palmer, M., Wils, J., Baron, B., Pignatti, F., Schoffski, P., Micheel, S., and Hecker, H. (2002). Clinical determinants of survival in patients with 5-fluorouracil-based treatment for metastatic colorectal cancer: results of a multivariate analysis of 3825 patients. *Ann Oncol*, 13(2):308–317.

Koopman, M., Antonini, N. F., Douma, J., Wals, J., Honkoop, A. H., Erdkamp, F. L., de Jong, R. S., Rodenburg, C. J., Vreugdenhil, G., Loosveld, O. J., van Bochove, A., Sinnige, H. A., Creemers, G. M., Tesselaar, M. E., Slee, P., Werter, M. J., Mol, L., Dalesio, O., and Punt, C. J. (2007). Sequential versus combination chemotherapy with capecitabine, irinotecan, and oxaliplatin in advanced colorectal cancer (CAIRO): a phase III randomised controlled trial. *Lancet*, 370(9582):135–142.

Kopetz, S., Desai, J., Chan, E., Hecht, J. R., O'Dwyer, P. J., Maru, D., Morris, V., Janku, F., Dasari, A., Chung, W., Issa, J.-P. J., Gibbs, P., James, B., Powis, G., Nolop, K. B., Bhattacharya, S., and Saltz, L. (2015). Phase II Pilot Study of Vemurafenib in Patients With Metastatic BRAF-Mutated Colorectal Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 33(34):4032–4038.

Kopetz, S., Grothey, A., Yaeger, R., Van Cutsem, E., Desai, J., Yoshino, T., Wasan, H., Ciardiello, F., Loupakis, F., Hong, Y. S., Steeghs, N., Guren, T. K., Arkenau, H.-T., Garcia-Alfonso, P., Pfeiffer, P., Orlov, S., Lonardi, S., Elez, E., Kim, T.-W., Schellens, J. H. M., Guo, C., Krishnan, A., Dekervel, J., Morris, V., Calvo Ferrandiz, A., Tarpgaard, L. S., Braun, M., Gollerkeri, A., Keir, C., Maharry, K., Pickard, M., Christy-Bittel, J., Anderson, L., Sandor, V., and Tabernero, J. (2019). Encorafenib, Binimetinib, and Cetuximab in BRAF V600E–Mutated Colorectal Cancer. *New England Journal of Medicine*, 381(17):1632–1643.

Kothari, N., Schell, M. J., Teer, J. K., Yeatman, T., Shibata, D., and Kim, R. (2014). Comparison of KRAS mutation analysis of colorectal cancer samples by standard testing and next-generation sequencing. *J Clin Pathol*, 67(9):764–767.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction.

Kuipers, E. J., Grady, W. M., Lieberman, D., Seufferlein, T., Sung, J. J., Boelens, P. G., van de Velde, C. J., and Watanabe, T. (2015). Colorectal cancer. *Nat Rev Dis Primers*, 1:15065.

Labianca, R., Nordlinger, B., Beretta, G. D., Mosconi, S., Mandala, M., Cervantes, A., and Arnold, D. (2013). Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 24 Suppl 6:vi64–72.

Lan, Q., Hsiung, C. A., Matsuo, K., Hong, Y. C., Seow, A., Wang, Z., Hosgood 3rd, H. D., Chen, K., Wang, J. C., Chatterjee, N., Hu, W., Wong, M. P., Zheng, W., Caporaso, N., Park, J. Y., Chen, C. J., Kim, Y. H., Kim, Y. T., Landi, M. T., Shen, H., Lawrence, C., Burdett, L., Yeager, M., Yuenger, J., Jacobs, K. B., Chang, I. S., Mitsudomi, T., Kim, H. N., Chang, G. C., Bassig, B. A., Tucker, M., Wei, F., Yin, Z., Wu, C., An, S. J., Qian, B., Lee, V. H., Lu, D., Liu, J., Jeon, H. S., Hsiao, C. F., Sung, J. S., Kim, J. H., Gao, Y. T., Tsai, Y. H., Jung, Y. J., Guo, H., Hu, Z., Hutchinson, A., Wang, W. C., Klein, R., Chung, C. C., Oh, I. J., Chen, K. Y., Berndt, S. I., He, X., Wu, W., Chang, J., Zhang, X. C., Huang, M. S., Zheng, H., Wang, J., Zhao, X., Li, Y., Choi, J. E., Su, W. C., Park, K. H., Sung, S. W., Shu, X. O., Chen, Y. M., Liu, L., Kang, C. H., Hu, L., Chen, C. H., Pao, W., Kim, Y. C., Yang, T. Y., Xu, J., Guan, P., Tan, W., Su, J., Wang, C. L., Li, H., Sihoe, A. D., Zhao, Z., Chen, Y., Choi, Y. Y., Hung, J. Y., Kim, J. S., Yoon, H. I., Cai, Q., Lin, C. C., Park, I. K., Xu, P., Dong, J., Kim, C., He, Q., Perng, R. P., Kohno, T., Kweon, S. S., Chen, C. Y., Vermeulen, R., Wu, J., Lim, W. Y., Chen, K. C., Chow, W. H., Ji, B. T., Chan, J. K., Chu, M., Li, Y. J., Yokota, J., Li, J., Chen, H., Xiang, Y. B., Yu, C. J., Kunitoh, H., Wu, G., Jin, L., Lo, Y. L., Shiraishi, K., Chen, Y. H., Lin, H. C., Wu, T., Wu, Y. L., Yang, P. C., Zhou, B., Shin, M. H., Fraumeni Jr., J. F., Lin, D., Chanock, S. J., and Rothman, N. (2012). Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat Genet*, 44(12):1330–1335.

Law, P. J., Timofeeva, M., Fernandez-Rozadilla, C., Broderick, P., Studd, J., Fernandez-Tajes, J., Farrington, S., Svinti, V., Palles, C., Orlando, G., Sud, A., Holroyd, A., Penegar, S., Theodoratou, E., Vaughan-Shaw, P., Campbell, H., Zgaga, L., Hayward, C., Campbell, A., Harris, S., Deary, I. J., Starr, J., Gatcombe, L., Pinna, M., Briggs, S., Martin, L., Jaeger, E., Sharma-Oates, A., East, J., Leedham, S., Arnold, R., Johnstone, E., Wang, H., Kerr, D., Kerr, R., Maughan, T., Kaplan, R., Al-Tassan, N., Palin, K., Hänninen, U. A., Cajuso, T., Tanskanen, T., Kondelin, J., Kaasinen, E., Sarin, A.-P., Eriksson, J. G., Rissanen, H., Knekt, P., Pukkala, E., Jousilahti, P., Salomaa, V., Ripatti, S., Palotie, A., Renkonen-Sinisalo, L., Lepistö, A., Böhm, J., Mecklin, J.-P., Buchanan, D. D., Win, A.-K., Hopper, J., Jenkins, M. E., Lindor, N. M., Newcomb, P. A., Gallinger, S., Duggan, D., Casey, G., Hoffmann, P., Nöthen, M. M., Jöckel, K.-H., Easton, D. F., Pharoah, P. D. P., Peto, J., Canzian, F., Swerdlow, A., Eeles, R. A., Kote-Jarai, Z., Muir, K., Pashayan, N., Henderson, B. E., Haiman, C. A., Schumacher, F. R., Al Olama, A. A., Benlloch, S., Berndt, S. I., Conti, D. V., Wiklund, F., Chanock, S., Gapstur, S., Stevens, V. L., Tangen, C. M., Batra, J., Clements, J., Gronberg, H., Schleutker, J., Albanes, D., Wolk, A., West, C., Mucci, L., Cancel-Tassin, G., Koutros, S., Sorensen, K. D., Grindedal, E. M., Neal, D. E., Hamdy, F. C., Donovan, J. L., Travis, R. C., Hamilton, R. J., Ingles, S. A., Rosenstein, B. S., Lu, Y.-J., Giles, G. G., Kibel, A. S., Vega, A., Kogevinas, M., Penney, K. L., Park, J. Y., Stanford, J. L., Cybulski, C., Nordestgaard, B. G., Maier, C., Kim, J., John, E. M., Teixeira, M. R., Neuhausen, S. L., De Ruyck, K., Razack, A., Newcomb, L. F., Gamulin, M., Kaneva, R., Usmani, N., Claessens, F., Townsend, P. A., Gago-Dominguez, M., Roobol, M. J., Menegaux, F., Khaw, K.-T., Cannon-Albright, L., Pandha, H., Thibodeau, S. N., Harkin, A., Allan, K., McQueen, J., Paul, J., Iveson, T., Saunders, M., Butterbach, K., Chang-Claude, J., Hoffmeister, M., Brenner, H., Kirac, I., Matošević, P., Hofer, P., Brezina, S., Gsur, A., Cheadle, J. P., Aaltonen, L. A., Tomlinson, I., Houlston, R. S., Dunlop, M. G., and consortium The, P. (2019). Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nature Communications*, 10(1):2154.

Lee, G. H., Malietzis, G., Askari, A., Bernardo, D., Al-Hassi, H. O., and Clark, S. K. (2015). Is right-sided colon cancer different to left-sided colorectal cancer? - a systematic review. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, 41(3):300–308.

Leitch, E. F., Chakrabarti, M., Crozier, J. E., McKee, R. F., Anderson, J. H., Horgan, P. G., and McMillan, D. C. (2007). Comparison of the prognostic value of selected markers of the systemic inflammatory response in patients with colorectal cancer. *Br J Cancer*, 97(9):1266–1270.

Lettre, G., Lange, C., and Hirschhorn, J. N. (2007). Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet Epidemiol*, 31(4):358–362.

Li, M., Li, C., and Guan, W. (2008). Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur J Hum Genet*, 16(5):635–643.

Li, N. (2016). Platelets in cancer metastasis: To help the ”villain” to do evil. *Int J Cancer*, 138(9):2078–2087.

Li, X., Ji, Y., Han, G., Li, X., Fan, Z., Li, Y., Zhong, Y., Cao, J., Zhao, J., Zhang, M., Wen, J., Goscinski, M. A., Nesland, J. M., and Suo, Z. (2016a). MPC1 and MPC2 expressions are associated with favorable clinical outcomes in prostate cancer. *BMC Cancer*, 16(1):894.

Li, X., Xun, Z., and Yang, Y. (2016b). Inhibition of phosphoserine phosphatase enhances the anticancer efficacy of 5-fluorouracil in colorectal cancer. *Biochem Biophys Res Commun*, 477(4):633–639.

Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. *Annu Rev Genomics Hum Genet*, 10:387–406.

Liang, P. S., Chen, T. Y., and Giovannucci, E. (2009). Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. *Int J Cancer*, 124(10):2406–2415.

Liao, L., Ge, M., Zhan, Q., Huang, R., Ji, X., Liang, X., and Zhou, X. (2019). PSPH Mediates the Metastasis and Proliferation of Non-small Cell Lung Cancer through MAPK Signaling Pathways. *Int J Biol Sci*, 15(1):183–194.

Lievre, A., Bachet, J. B., Le Corre, D., Boige, V., Landi, B., Emile, J. F., Cote, J. F., Tomasic, G., Penna, C., Ducreux, M., Rougier, P., Penault-Llorca, F., and Laurent-Puig, P. (2006). KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res*, 66(8):3992–3995.

Lin, P.-C., Yeh, Y.-M., Wu, P.-Y., Hsu, K.-F., Chang, J.-Y., and Shen, M.-R. (2019). Germline susceptibility variants impact clinical outcome and therapeutic strategies for stage III colorectal cancer. *Scientific reports*, 9(1):3931.

Liou, J. M., Shun, C. T., Liang, J. T., Chiu, H. M., Chen, M. J., Chen, C. C., Wang, H. P., Wu, M. S., and Lin, J. T. (2010). Plasma insulin-like growth factor-binding protein-2 levels as diagnostic and prognostic biomarker of colorectal cancer. *J Clin Endocrinol Metab*, 95(4):1717–1725.

Lochhead, P., Kuchiba, A., Imamura, Y., Liao, X., Yamauchi, M., Nishihara, R., Qian, Z. R., Morikawa, T., Shen, J., Meyerhardt, J. A., Fuchs, C. S., and Ogino, S. (2013). Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. *J Natl Cancer Inst*, 105(15):1151–1156.

Loo, L. W., Cheng, I., Tiirikainen, M., Lum-Jones, A., Seifried, A., Dunklee, L. M., Church, J. M., Gryfe, R., Weisenberger, D. J., Haile, R. W., Gallinger, S., Duggan, D. J., Thibodeau, S. N., Casey, G., and Le Marchand, L. (2012). cis-Expression QTL analysis of established colorectal cancer risk variants in colon tumors and adjacent normal tissue. *PLoS One*, 7(2):e30477.

Loo, L. W. M., Lemire, M., and Le Marchand, L. (2017). In silico pathway analysis and tissue specific cis-eQTL for colorectal cancer GWAS risk variants. *BMC Genomics*, 18(1):381.

Lynch, H. T. and de la Chapelle, A. (2003). Hereditary colorectal cancer. *N Engl J Med*, 348(10):919–932.

Madi, A., Fisher, D., Wilson, R. H., Adams, R. A., Meade, A. M., Kenny, S. L., Nichols, L. L., Seymour, M. T., Wasan, H., Kaplan, R., and Maughan, T. S. (2012). Oxaliplatin/capecitabine vs oxaliplatin/infusional 5-FU in advanced colorectal cancer: the MRC COIN trial. *Br J Cancer*, 107(7):1037–1043.

Majek, O., Gondos, A., Jansen, L., Emrich, K., Holleczek, B., Katalinic, A., Nennecke, A., Eberle, A., and Brenner, H. (2013). Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in Germany. *PLoS One*, 8(7):e68077.

Malumbres, M. and Barbacid, M. (2003). RAS oncogenes: the first 30 years. *Nat Rev Cancer*, 3(6):459–465.

Manolio, T. A. (2010). Genomewide association studies and assessment of the risk of disease. *N Engl J Med*, 363(2):166–176.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A., and Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265):747–753.

Manson, J. E., Buring, J. E., Satterfield, S., and Hennekens, C. H. (1991). Baseline characteristics of participants in the Physicians' Health Study: a randomized trial of aspirin and beta-carotene in U.S. physicians. *American journal of preventive medicine*, 7(3):150–154.

Marchini, J. and Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nat Rev Genet*, 11(7):499–511.

Marcuello, E., Altes, A., del Rio, E., Cesar, A., Menoyo, A., and Baiget, M. (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. *Int J Cancer*, 112(5):733–737.

Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R., and Campbell, P. J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*, 171(5):1029–1041.e21.

Maughan, T. S., Adams, R. A., Smith, C. G., Meade, A. M., Seymour, M. T., Wilson, R. H., Idziaszczyk, S., Harris, R., Fisher, D., Kenny, S. L., Kay, E., Mitchell, J. K., Madi, A., Jasani, B., James, M. D., Bridgewater, J., Kennedy, M. J., Claes, B., Lambrechts, D., Kaplan, R., and Cheadle, J. P. (2011). Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet*, 377(9783):2103–2114.

Maughan, T. S., Meade, A. M., Adams, R. A., Richman, S. D., Butler, R., Fisher, D., Wilson, R. H., Jasani, B., Taylor, G. R., Williams, G. T., Sampson, J. R., Seymour, M. T., Nichols, L. L., Kenny, S. L., Nelson, A., Sampson, C. M., Hodgkinson, E., Bridgewater, J. A., Furniss, D. L., Roy, R., Pope, M. J., Pope, J. K., Parmar, M., Quirke, P., and Kaplan, R. (2014). A feasibility study testing four hypotheses with phase II outcomes in advanced colorectal cancer (MRC FOCUS3): a model for randomised controlled trials in the era of personalised medicine? *Br J Cancer*, 110(9):2178–2186.

McFarland, C. D., Yaglom, J. A., Wojtkowiak, J. W., Scott, J. G., Morse, D. L., Sherman, M. Y., and Mirny, L. A. (2017). The Damaging Effect of Passenger Mutations on Cancer Progression. *Cancer Res*, 77(18):4763–4772.

McShane, L. M., Altman, D. G., Sauerbrei, W., Taube, S. E., Gion, M., and Clark, G. M. (2005). REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*, 93(4):387–391.

Mekenkamp, L. J., Heesterbeek, K. J., Koopman, M., Tol, J., Teerenstra, S., Venderbosch, S., Punt, C. J., and Nagtegaal, I. D. (2012). Mucinous adenocarcinomas: poor prognosis in metastatic colorectal cancer. *Eur J Cancer*, 48(4):501–509.

Mendelsohn, J. and Baselga, J. (2006). Epidermal growth factor receptor targeting in cancer. *Semin Oncol*, 33(4):369–385.

Midgley, R. S., McConkey, C. C., Johnstone, E. C., Dunn, J. A., Smith, J. L., Grumett, S. A., Julier, P., Iveson, C., Yanagisawa, Y., Warren, B., Langman, M. J., and Kerr, D. J. (2010). Phase III randomized trial assessing rofecoxib in the adjuvant setting of colorectal cancer: final results of the VICTOR trial. *J Clin Oncol*, 28(30):4575–4580.

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., Jemal, A., Kramer, J. L., and Siegel, R. L. (2019). Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin*, 69(5):363–385.

Missiaglia, E., Jacobs, B., D'Ario, G., Di Narzo, A. F., Soneson, C., Budinska, E., Popovici, V., Vecchione, L., Gerster, S., Yan, P., Roth, A. D., Klingbiel, D., Bosman, F. T., Delorenzi, M., and Tejpar, S. (2014). Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Annals of oncology : official journal of the European Society for Medical Oncology*, 25(10):1995–2001.

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

Morris, E. J., Penegar, S., Whiffin, N., Broderick, P., Bishop, D. T., Northwood, E., Quirke, P., Finan, P., and Houlston, R. S. (2015). A retrospective observational study of the relationship between single nucleotide polymorphisms associated with the risk of developing colorectal cancer and survival. *PLoS One*, 10(2):e0117816.

Mullard, A. (2019). Cracking KRAS.

Munro, A. J., Lain, S., and Lane, D. P. (2005). P53 abnormalities and outcomes in colorectal cancer: a systematic review. *British journal of cancer*, 92(3):434–444.

Nakamura, M., Yamada, Y., Muro, K., Takahashi, K., Baba, H., Sasaki, Y., Komatsu, Y., Satoh, T., Mishima, H., Watanabe, M., Sakata, Y., Morita, S., Shimada, Y., and Sugihara, K. (2015). The SOFT trial: a Phase III study of the dihydropyrimidine dehydrogenase inhibitory fluoropyrimidine S-1 and oxaliplatin (SOX) plus bevacizumab as first-line chemotherapy for metastatic colorectal cancer. *Future Oncol*, 11(10):1471–1478.

Nakazato, H., Takeshima, H., Kishino, T., Kubo, E., Hattori, N., Nakajima, T., Yamashita, S., Igaki, H., Tachimori, Y., Kuniyoshi, Y., and Ushijima, T. (2016). Early-Stage Induction of SWI/SNF Mutations during Esophageal Squamous Cell Carcinogenesis. *PLoS One*, 11(1):e0147372.

Negrini, S., Gorgoulis, V. G., and Halazonetis, T. D. (2010). Genomic instability–an evolving hallmark of cancer. *Nat Rev Mol Cell Biol*, 11(3):220–228.

Nica, A. C. and Dermitzakis, E. T. (2013). Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci*, 368(1620):20120362.

Nica, A. C., Montgomery, S. B., Dimas, A. S., Stranger, B. E., Beazley, C., Barroso, I., and Dermitzakis, E. T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet*, 6(4):e1000895.

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., and Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*, 6(4):e1000888.

Oetting, W. S., Jacobson, P. A., and Israni, A. K. (2017). Validation Is Critical for Genome-Wide Association Study-Based Associations. *Am J Transplant*, 17(2):318–319.

Ogino, S., Nosho, K., Kirkner, G. J., Kawasaki, T., Meyerhardt, J. A., Loda, M., Giovannucci, E. L., and Fuchs, C. S. (2009). CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut*, 58(1):90–96.

Ogunbiyi, O. A., Goodfellow, P. J., Herfarth, K., Gagliardi, G., Swanson, P. E., Birnbaum, E. H., Read, T. E., Fleshman, J. W., Kodner, I. J., and Moley, J. F. (1998). Confirmation that chromosome 18q allelic loss in colon cancer is a prognostic indicator. *J Clin Oncol*, 16(2):427–433.

Oliveira, A. F., Bretes, L., and Furtado, I. (2019). Review of PD-1/PD-L1 Inhibitors in Metastatic dMMR/MSI-H Colorectal Cancer. *Frontiers in oncology*, 9:396.

Pai, R. K., Jayachandran, P., Koong, A. C., Chang, D. T., Kwok, S., Ma, L., Arber, D. A., Balise, R. R., Tubbs, R. R., Shadrach, B., and Pai, R. K. (2012). BRAF-mutated, microsatellite-stable adenocarcinoma of the proximal colon: an aggressive adenocarcinoma with poor survival, mucinous differentiation, and adverse morphologic features. *The American journal of surgical pathology*, 36(5):744–752.

Palles, C., Cazier, J. B., Howarth, K. M., Domingo, E., Jones, A. M., Broderick, P., Kemp, Z., Spain, S. L., Guarino, E., Salguero, I., Sherborne, A., Chubb, D., Carvajal-Carmona, L. G., Ma, Y., Kaur, K., Dobbins, S., Barclay, E., Gorman, M., Martin, L., Kovac, M. B., Humphray, S., Lucassen, A., Holmes, C. C., Bentley, D., Donnelly, P., Taylor, J., Petridis, C., Roylance, R., Sawyer, E. J., Kerr, D. J., Clark, S., Grimes, J., Kearsey, S. E., Thomas, H. J., McVean, G., Houlston, R. S., and Tomlinson, I. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet*, 45(2):136–144.

Panagiotou, O. A., Evangelou, E., and Ioannidis, J. P. (2010). Genome-wide significant associations for variants with minor allele frequency of 5% or less–an overview: A HuGE review. *Am J Epidemiol*, 172(8):869–889.

Park, S. M., Seo, E. H., Bae, D. H., Kim, S. S., Kim, J., Lin, W., Kim, K. H., Park, J. B., Kim, Y. S., Yin, J., and Kim, S. Y. (2019). Phosphoserine Phosphatase Promotes Lung Cancer Progression through the Dephosphorylation of IRS-1 and a Noncanonical L-Serine-Independent Pathway. *Mol Cells*, 42(8):604–616.

Passarelli, M. N., Coghill, A. E., Hutter, C. M., Zheng, Y., Makar, K. W., Potter, J. D., and Newcomb, P. A. (2011). Common colorectal cancer risk variants in SMAD7 are associated with survival among prediagnostic nonsteroidal anti-inflammatory drug users: a population-based study of postmenopausal women. *Genes Chromosomes Cancer*, 50(11):875–886.

Peltomaki, P. (2001). Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum Mol Genet*, 10(7):735–740.

Peltomaki, P. (2003). Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J Clin Oncol*, 21(6):1174–1179.

Pendlebury, S., Duchesne, F., Reed, K. A., Smith, J. L., and Kerr, D. J. (2003). A trial of adjuvant therapy in colorectal cancer: the VICTOR trial. *Clin Colorectal Cancer*, 3(1):58–60.

Pereira, A. A. L., Rego, J. F. M., Morris, V., Overman, M. J., Eng, C., Garrett, C. R., Boutin, A. T., Ferrarotto, R., Lee, M., Jiang, Z.-Q., Hoff, P. M., Vauthey, J.-N., Vilar, E., Maru, D., and Kopetz, S. (2015). Association between KRAS mutation and lung metastasis in advanced colorectal cancer. *British journal of cancer*, 112(3):424–428.

Peters, U., Hutter, C. M., Hsu, L., Schumacher, F. R., Conti, D. V., Carlson, C. S., Edlund, C. K., Haile, R. W., Gallinger, S., Zanke, B. W., Lemire, M., Rangrej, J., Vijayaraghavan, R., Chan, A. T., Hazra, A., Hunter, D. J., Ma, J., Fuchs, C. S., Giovannucci, E. L., Kraft, P., Liu, Y., Chen, L., Jiao, S., Makar, K. W., Taverna, D., Gruber, S. B., Rennert, G., Moreno, V., Ulrich, C. M., Woods, M. O., Green, R. C., Parfrey, P. S., Prentice, R. L., Kooperberg, C., Jackson, R. D., Lacroix, A. Z., Caan, B. J., Hayes, R. B., Berndt, S. I., Chanock, S. J., Schoen, R. E., Chang-Claude, J., Hoffmeister, M., Brenner, H., Frank, B., Bezieau, S., Kury, S., Slattery, M. L., Hopper, J. L., Jenkins, M. A., Le Marchand, L., Lindor, N. M., Newcomb, P. A., Seminara, D., Hudson, T. J., Duggan, D. J., Potter, J. D., and Casey, G. (2012). Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet*, 131(2):217–234.

Peters, U., Jiao, S., Schumacher, F. R., Hutter, C. M., Aragaki, A. K., Baron, J. A., Berndt, S. I., Bezieau, S., Brenner, H., Butterbach, K., Caan, B. J., Campbell, P. T., Carlson, C. S., Casey, G., Chan, A. T., Chang-Claude, J., Chanock, S. J., Chen, L. S., Coetzee, G. A., Coetzee, S. G., Conti, D. V., Curtis, K. R., Duggan, D., Edwards, T., Fuchs, C. S., Gallinger, S., Giovannucci, E. L., Gogarten, S. M., Gruber, S. B., Haile, R. W., Harrison, T. A., Hayes, R. B., Henderson, B. E., Hoffmeister, M., Hopper, J. L., Hudson, T. J., Hunter, D. J., Jackson, R. D., Jee, S. H., Jenkins, M. A., Jia, W. H., Kolonel, L. N., Kooperberg, C., Kury, S., Lacroix,

A. Z., Laurie, C. C., Laurie, C. A., Le Marchand, L., Lemire, M., Levine, D., Lindor, N. M., Liu, Y., Ma, J., Makar, K. W., Matsuo, K., Newcomb, P. A., Potter, J. D., Prentice, R. L., Qu, C., Rohan, T., Rosse, S. A., Schoen, R. E., Seminara, D., Shrubsole, M., Shu, X. O., Slattery, M. L., Taverna, D., Thibodeau, S. N., Ulrich, C. M., White, E., Xiang, Y., Zanke, B. W., Zeng, Y. X., Zhang, B., Zheng, W., and Hsu, L. (2013). Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology*, 144(4):799–807.e24.

Phipps, A. I., Buchanan, D. D., Makar, K. W., Burnett-Hartman, A. N., Coghill, A. E., Passarelli, M. N., Baron, J. A., Ahnen, D. J., Win, A. K., Potter, J. D., and Newcomb, P. A. (2012). BRAF mutation status and survival after colorectal cancer diagnosis according to patient and tumor characteristics. *Cancer Epidemiol Biomarkers Prev*, 21(10):1792–1798.

Phipps, A. I., Buchanan, D. D., Makar, K. W., Win, A. K., Baron, J. A., Lindor, N. M., Potter, J. D., and Newcomb, P. A. (2013). KRAS-mutation status in relation to colorectal cancer survival: the joint impact of correlated tumour markers. *Br J Cancer*, 108(8):1757–1764.

Phipps, A. I., Passarelli, M. N., Chan, A. T., Harrison, T. A., Jeon, J., Hutter, C. M., Berndt, S. I., Brenner, H., Caan, B. J., Campbell, P. T., Chang-Claude, J., Chanock, S. J., Cheadle, J. P., Curtis, K. R., Duggan, D., Fisher, D., Fuchs, C. S., Gala, M., Giovannucci, E. L., Hayes, R. B., Hoffmeister, M., Hsu, L., Jacobs, E. J., Jansen, L., Kaplan, R., Kap, E. J., Maughan, T. S., Potter, J. D., Schoen, R. E., Seminara, D., Slattery, M. L., West, H., White, E., Peters, U., and Newcomb, P. A. (2016). Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *Carcinogenesis*, 37(1):87–95.

Pink, R. C., Wicks, K., Caley, D. P., Punch, E. K., Jacobs, L., and Carter, D. R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? *Rna*, 17(5):792–798.

Pongpanich, M., Sullivan, P. F., and Tzeng, J. Y. (2010). A quality control algorithm for filtering SNPs in genome-wide association studies. *Bioinformatics*, 26(14):1731–1737.

Poole, L. A. and Cortez, D. (2017). Functions of SMARCAL1, ZRANB3, and HLTF in maintaining genome stability. *Crit Rev Biochem Mol Biol*, 52(6):696–714.

Popat, S., Hubner, R., and Houlston, R. S. (2005). Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol*, 23(3):609–618.

Porcu, E., Rueger, S., Lepik, K., Santoni, F. A., Reymond, A., and Kutalik, Z. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun*, 10(1):3300.

Porru, M., Pompili, L., Caruso, C., Biroccio, A., and Leonetti, C. (2018). Targeting KRAS in metastatic colorectal cancer: current strategies and emerging opportunities. *Journal of experimental & clinical cancer research : CR*, 37(1):57.

Potter, J. D. (1999). Colorectal cancer: molecules and populations. *J Natl Cancer Inst*, 91(11):916–932.

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

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., and Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, 26(18):2336–2337.

Puccetti, M. V., Adams, C. M., Kushinsky, S., and Eischen, C. M. (2019). Smarcal1 and Zranb3 Protect Replication Forks from Myc-Induced DNA Replication Stress. *Cancer Res*, 79(7):1612–1623.

Pylayeva-Gupta, Y., Grabocka, E., and Bar-Sagi, D. (2011). RAS oncogenes: weaving a tumorigenic web. *Nat Rev Cancer*, 11(11):761–774.

Rajagopalan, H., Bardelli, A., Lengauer, C., Kinzler, K. W., Vogelstein, B., and Velculescu, V. E. (2002). Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature*, 418(6901):934.

Raychaudhuri, S. (2011). Mapping rare and common causal alleles for complex human diseases. *Cell*, 147(1):57–69.

Reed, E., Nunez, S., Kulp, D., Qian, J., Reilly, M. P., and Foulkes, A. S. (2015). A guide to genome-wide association analysis and post-analytic interrogation. *Stat Med*, 34(28):3769–3792.

Reeves, G. K., Pirie, K., Beral, V., Green, J., Spencer, E., and Bull, D. (2007). Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ*, 335(7630):1134.

Richman, S. D., Seymour, M. T., Chambers, P., Elliott, F., Daly, C. L., Meade, A. M., Taylor, G., Barrett, J. H., and Quirke, P. (2009). KRAS and BRAF mutations in advanced colorectal cancer are associated with poor prognosis but do not preclude benefit from oxaliplatin or irinotecan: results from the MRC FOCUS trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(35):5931–5937.

Riihimaki, M., Hemminki, A., Sundquist, J., and Hemminki, K. (2016). Patterns of metastasis in colon and rectal cancer. *Sci Rep*, 6:29765.

Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281):1516–1517.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature*, 405(6788):847–856.

Roberts, P. J. and Stinchcombe, T. E. (2013). KRAS mutation: should we test for it, and does it matter? *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 31(8):1112–1121.

Robles-Zurita, J., Boyd, K. A., Briggs, A. H., Iveson, T., Kerr, R. S., Saunders, M. P., Cassidy, J., Hollander, N. H., Tabernero, J., Segelov, E., Glimelius, B., Harkin, A., Allan, K., McQueen, J., Pearson, S., Waterston, A., Medley, L., Wilson, C., Ellis, R., Essapen, S., Dhadda, A. S., Hughes, R., Falk, S., Raouf, S., Rees, C., Olesen, R. K., Propper, D., Bridgewater, J., Azzabi, A., Farrugia, D., Webb, A., Cunningham, D., Hickish, T., Weaver, A., Gollins, S., Wasan, H. S., and Paul, J. (2018). SCOT: a comparison of cost-effectiveness from a large randomised phase III trial of two durations of adjuvant Oxaliplatin combination chemotherapy for colorectal cancer. *British journal of cancer*, 119(11):1332–1338.

Romagnoni, A., Jegou, S., Van Steen, K., Wainrib, G., and Hugot, J.-P. (2019). Comparative performances of machine learning methods for classifying Crohn Disease patients using genome-wide genotyping data. *Scientific reports*, 9(1):10351.

Rosmarin, D., Palles, C., Church, D., Domingo, E., Jones, A., Johnstone, E., Wang, H., Love, S., Julier, P., Scudder, C., Nicholson, G., Gonzalez-Neira, A., Martin, M., Sargent, D., Green, E., McLeod, H., Zanger, U. M., Schwab, M., Braun, M., Seymour, M., Thompson, L., Lacas, B., Boige, V., Ribelles, N., Afzal, S., Enghusen, H., Jensen, S. A., Etienne-Grimaldi, M. C., Milano, G., Wadelius, M., Glimelius, B., Garmo, H., Gusella, M., Lecomte, T., Laurent-Puig, P., Martinez-Balibrea, E., Sharma, R., Garcia-Foncillas, J., Kleibl, Z., Morel, A., Pignon, J. P., Midgley, R., Kerr, D., and Tomlinson, I. (2014). Genetic markers of toxicity from capecitabine and other fluorouracil-based regimens: investigation in the QUASAR2 study, systematic review, and meta-analysis. *J Clin Oncol*, 32(10):1031–1039.

Rossi, T., Pistoni, M., Sancisi, V., Gobbi, G., Torricelli, F., Donati, B., Ribisi, S., Gugnoni, M., and Ciarrocchi, A. (2019). RAIN is a novel Enhancer-associated lncRNA that controls RUNX2 expression and promotes breast and thyroid cancer. *Mol Cancer Res*.

Rosty, C., Young, J. P., Walsh, M. D., Clendenning, M., Walters, R. J., Pearson, S., Pavluk, E., Nagler, B., Pakenas, D., Jass, J. R., Jenkins, M. A., Win, A. K., Southey, M. C., Parry, S., Hopper, J. L., Giles, G. G., Williamson, E., English, D. R., and Buchanan, D. D. (2013). Colorectal carcinomas with KRAS mutation are associated with distinctive morphological and molecular features. *Mod Pathol*, 26(6):825–834.

Russo, A., Bazan, V., Iacopetta, B., Kerr, D., Soussi, T., and Gebbia, N. (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. *J Clin Oncol*, 23(30):7518–7528.

Sadanandam, A., Lyssiotis, C. A., Homicsko, K., Collisson, E. A., Gibb, W. J., Wullschleger, S., Ostos, L. C., Lannon, W. A., Grotzinger, C., Del Rio, M., Lhermitte, B., Olshen, A. B., Wiedenmann, B., Cantley, L. C., Gray, J. W., and Hanahan, D. (2013). A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med*, 19(5):619–625.

Samowitz, W. S., Sweeney, C., Herrick, J., Albertsen, H., Levin, T. R., Murtaugh, M. A., Wolff, R. K., and Slattery, M. L. (2005). Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers. *Cancer Res*, 65(14):6063–6069.

Sandoval, I. T., Delacruz, R. G., Miller, B. N., Hill, S., Olson, K. A., Gabriel, A. E., Boyd, K., Satterfield, C., Van Remmen, H., Rutter, J., and Jones, D. A. (2017). A metabolic switch controls intestinal differentiation downstream of Adenomatous polyposis coli (APC). *Elife*, 6.

Sanoff, H. K., Renfro, L. A., Poonnen, P., Ambadwar, P., Sargent, D. J., Goldberg, R. M., and McLeod, H. (2014). Germline variation in colorectal risk Loci does not influence treatment effect or survival in metastatic colorectal cancer. *PLoS One*, 9(4):e94727.

Sansregret, L., Vanhaesebroeck, B., and Swanton, C. (2018). Determinants and clinical implications of chromosomal instability in cancer. *Nat Rev Clin Oncol*, 15(3):139–150.

Sato, K., Masuda, T., Hu, Q., Tobo, T., Kidogami, S., Ogawa, Y., Saito, T., Nambara, S., Komatsu, H., Hirata, H., Sakimura, S., Uchi, R., Hayashi, N., Iguchi, T., Eguchi, H., Ito, S., Nakagawa, T., and Mimori, K. (2017). Phosphoserine Phosphatase Is a Novel Prognostic Biomarker on Chromosome 7 in Colorectal Cancer. *Anticancer Res*, 37(5):2365–2371.

Schell, J. C., Olson, K. A., Jiang, L., Hawkins, A. J., Van Vranken, J. G., Xie, J., Egnatchik, R. A., Earl, E. G., DeBerardinis, R. J., and Rutter, J. (2014). A role for the mitochondrial pyruvate carrier as a repressor of the Warburg effect and colon cancer cell growth. *Mol Cell*, 56(3):400–413.

Schell, J. C., Wisidagama, D. R., Bensard, C., Zhao, H., Wei, P., Tanner, J., Flores, A., Mohlman, J., Sorensen, L. K., Earl, C. S., Olson, K. A., Miao, R., Waller, T. C., Delker, D., Kanth, P., Jiang, L., DeBerardinis, R. J., Bronner, M. P., Li, D. Y., Cox, J. E., Christofk, H. R., Lowry, W. E., Thummel, C. S., and Rutter, J. (2017). Control of intestinal stem cell function and proliferation by mitochondrial pyruvate metabolism. *Nat Cell Biol*, 19(9):1027–1036.

Schirripa, M., Cremolini, C., Loupakis, F., Morvillo, M., Bergamo, F., Zoratto, F., Salvatore, L., Antoniotti, C., Marmorino, F., Sensi, E., Lupi, C., Fontanini, G., De Gregorio, V., Giannini, R., Basolo, F., Masi, G., and Falcone, A. (2015). Role of NRAS mutations as prognostic and predictive markers in metastatic colorectal cancer. *Int J Cancer*, 136(1):83–90.

Schmoll, H. J., Arnold, D., de Gramont, A., Ducreux, M., Grothey, A., O'Dwyer, P. J., Van Cutsem, E., Hermann, F., Bosanac, I., Bendahmane, B., Mancao, C., and Tabernero, J. (2018). MODUL-a multicenter randomized clinical trial of biomarker-driven maintenance therapy following first-line standard induction treatment of metastatic colorectal cancer: an adaptable signal-seeking approach. *J Cancer Res Clin Oncol*, 144(6):1197–1204.

Sedgwick, P. and Marston, L. (2015). How to read a funnel plot in a meta-analysis. *BMJ (Clinical research ed.)*, 351:h4718.

Seymour, M. T., Brown, S. R., Middleton, G., Maughan, T., Richman, S., Gwyther, S., Lowe, C., Seligmann, J. F., Wadsley, J., Maisey, N., Chau, I., Hill, M., Dawson, L., Falk, S., O'Callaghan, A., Benstead, K., Chambers, P., Oliver, A., Marshall, H., Napp, V., and Quirke, P. (2013). Panitumumab and irinotecan versus irinotecan alone for patients with KRAS wild-type, fluorouracil-resistant advanced colorectal cancer (PICCOLO): a prospectively stratified randomised trial. *Lancet Oncol*, 14(8):749–759.

Seymour, M. T., Maughan, T. S., Ledermann, J. A., Topham, C., James, R., Gwyther, S. J., Smith, D. B., Shepherd, S., Maraveyas, A., Ferry, D. R., Meade, A. M., Thompson, L., Griffiths, G. O., Parmar, M. K., and Stephens, R. J. (2007). Different strategies of sequential and combination chemotherapy for patients with poor prognosis advanced colorectal cancer (MRC FOCUS): a randomised controlled trial. *Lancet*, 370(9582):143–152.

Seymour, M. T., Thompson, L. C., Wasan, H. S., Middleton, G., Brewster, A. E., Shepherd, S. F., O'Mahony, M. S., Maughan, T. S., Parmar, M., and Langley, R. E. (2011). Chemotherapy options in elderly and frail patients with metastatic colorectal cancer (MRC FOCUS2): an open-label, randomised factorial trial. *Lancet*, 377(9779):1749–1759.

Sham, P. C. and Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet*, 15(5):335–346.

Shastry, B. S. (2009). SNPs: impact on gene function and phenotype. *Methods Mol Biol*, 578:3–22.

Shi, Y., Hu, Z., Wu, C., Dai, J., Li, H., Dong, J., Wang, M., Miao, X., Zhou, Y., Lu, F., Zhang, H., Hu, L., Jiang, Y., Li, Z., Chu, M., Ma, H., Chen, J., Jin, G., Tan, W., Wu, T., Zhang, Z., Lin, D., and Shen, H. (2011). A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet*, 43(12):1215–1218.

Shirasawa, S., Furuse, M., Yokoyama, N., and Sasazuki, T. (1993). Altered growth of human colon cancer cell lines disrupted at activated Ki-ras. *Science (New York, N.Y.)*, 260(5104):85–88.

Simkens, L. H., van Tinteren, H., May, A., ten Tije, A. J., Creemers, G. J., Loosveld, O. J., de Jongh, F. E., Erdkamp, F. L., Erjavec, Z., van der Torren, A. M., Tol, J., Braun, H. J., Nieboer, P., van der Hoeven, J. J., Haasjes, J. G., Jansen, R. L., Wals, J., Cats, A., Derleyn, V. A., Honkoop, A. H., Mol, L., Punt, C. J., and Koopman, M. (2015). Maintenance treatment with capecitabine and bevacizumab in metastatic colorectal cancer (CAIRO3): a phase 3 randomised controlled trial of the Dutch Colorectal Cancer Group. *Lancet*, 385(9980):1843–1852.

Sinicrope, F. A. and Sargent, D. J. (2012). Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications. *Clin Cancer Res*, 18(6):1506–1512.

Slatkin, M. (2008). Linkage disequilibrium–understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*, 9(6):477–485.

Slattery, M. L. (2000). Diet, lifestyle, and colon cancer. *Semin Gastrointest Dis*, 11(3):142–146.

Smalley, K. S. M., Xiao, M., Villanueva, J., Nguyen, T. K., Flaherty, K. T., Letrero, R., Van Belle, P., Elder, D. E., Wang, Y., Nathanson, K. L., and Herlyn, M. (2009). CRAF inhibition induces apoptosis in melanoma cells with non-V600E BRAF mutations. *Oncogene*, 28(1):85–94.

Smith, C. G., Fisher, D., Claes, B., Maughan, T. S., Idziaszczyk, S., Peuteman, G., Harris, R., James, M. D., Meade, A., Jasani, B., Adams, R. A., Kenny, S., Kaplan, R., Lambrechts, D., and Cheadle, J. P. (2013). Somatic profiling of the epidermal growth factor receptor pathway in tumors from patients with advanced colorectal cancer treated with chemotherapy +/- cetuximab. *Clin Cancer Res*, 19(15):4104–4113.

Smith, C. G., Fisher, D., Harris, R., Maughan, T. S., Phipps, A. I., Richman, S. D., Seymour, M. T., Tomlinson, I. P., Rosmarin, D., Kerr, D. J., Chan, A. T., Peters, U., Newcomb, P. A., Idziaszczyk, S., West, H., Meade, A., Kaplan, R., and Cheadle, J. P. (2015). Analyses of 7,635 patients with colorectal cancer using independent training and validation cohorts show that rs9929218 in CDH1 is a prognostic marker of survival. *Clin Cancer Res*.

Stark, A. E. (2015). Estimation of divergence from Hardy-Weinberg form. *Twin Res Hum Genet*, 18(4):399–405.

Steering Committee of the Physicians' Health Study Research Group (1989). Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med*, 321(3):129–135.

Steinberg, S. M., Barkin, J. S., Kaplan, R. S., and Stablein, D. M. (1986). Prognostic indicators of colon tumors. The Gastrointestinal Tumor Study Group experience. *Cancer*, 57(9):1866–1870.

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rucker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., and Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Clinical research ed.)*, 343:d4002.

Stinchcombe, T. E. and Der, C. J. (2011). Are all KRAS mutations created equal? *The Lancet. Oncology*, 12(8):717–718.

Stolze, B., Reinhart, S., Bulllinger, L., Frohling, S., and Scholl, C. (2015). Comparative analysis of KRAS codon 12, 13, 18, 61, and 117 mutations using human MCF10A isogenic cell lines. *Scientific reports*, 5:8535.

Stratton, M. R., Campbell, P. J., and Futreal, P. A. (2009). The cancer genome. *Nature*, 458(7239):719–724.

Swanton, C., Tomlinson, I., and Downward, J. (2006). Chromosomal instability, colorectal cancer and taxane resistance. *Cell Cycle*, 5(8):818–823.

Szymczak, S., Biernacka, J. M., Cordell, H. J., Gonzalez-Recio, O., Konig, I. R., Zhang, H., and Sun, Y. V. (2009). Machine learning in genome-wide association studies.

t Lam-Boer, J., Mol, L., Verhoef, C., de Haan, A. F., Yilmaz, M., Punt, C. J., de Wilt, J. H., and Koopman, M. (2014). The CAIRO4 study: the role of surgery of the primary tumour with few or absent symptoms in patients with synchronous unresectable metastases of colorectal cancer–a randomized phase III study of the Dutch Colorectal Cancer Group (DCCG). *BMC Cancer*, 14:741.

Taieb, J., Lapeyre-Prost, A., Laurent Puig, P., and Zaanan, A. (2019). Exploring the best treatment options for BRAF-mutant metastatic colon cancer. *British Journal of Cancer*, 121(6):434–442.

Takahashi, Y., Sawada, G., Kurashige, J., Uchi, R., Matsumura, T., Ueo, H., Takano, Y., Eguchi, H., Sudo, T., Sugimachi, K., Yamamoto, H., Doki, Y., Mori, M., and Mimori, K. (2014). Amplification of PVT-1 is involved in poor prognosis via apoptosis inhibition in colorectal cancers. *Br J Cancer*, 110(1):164–171.

Takatsuno, Y., Mimori, K., Yamamoto, K., Sato, T., Niida, A., Inoue, H., Imoto, S., Kawano, S., Yamaguchi, R., Toh, H., Iinuma, H., Ishimaru, S., Ishii, H., Suzuki, S., Tokudome, S., Watanabe, M., Tanaka, J., Kudo, S. E., Mochizuki, H., Kusunoki, M., Yamada, K., Shimada, Y., Moriya, Y., Miyano, S., Sugihara, K., and Mori, M. (2013). The rs6983267 SNP is associated with MYC transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. *Ann Surg Oncol*, 20(4):1395–1402.

Tang, X. P., Chen, Q., Li, Y., Wang, Y., Zou, H. B., Fu, W. J., Niu, Q., Pan, Q. G., Jiang, P., Xu, X. S., Zhang, K. Q., Liu, H., Bian, X. W., and Wu, X. F. (2019). Mitochondrial pyruvate carrier 1 functions as a tumor suppressor and predicts the prognosis of human renal cell carcinoma. *Lab Invest*, 99(2):191–199.

Temko, D., Tomlinson, I. P. M., Severini, S., Schuster-Bockler, B., and Graham, T. A. (2018). The effects of mutational processes and selection on driver mutations across cancer types. *Nat Commun*, 9(1):1857.

Tenesa, A., Theodoratou, E., Din, F. V., Farrington, S. M., Cetnarskyj, R., Barnetson, R. A., Porteous, M. E., Campbell, H., and Dunlop, M. G. (2010). Ten common genetic variants associated with colorectal cancer risk are not associated with survival after diagnosis. *Clin Cancer Res*, 16(14):3754–3759.

The Women's Health Initiative Study Group (1998). Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials*, 19(1):61–109.

Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., Crenshaw, A., Cancel-Tassin, G., Staats, B. J., Wang, Z., Gonzalez-Bosquet, J., Fang, J., Deng, X., Berndt, S. I., Calle, E. E., Feigelson, H. S., Thun, M. J., Rodriguez, C., Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F. R., Giovannucci, E., Willett, W. C., Cussenot, O., Valeri, A., Andriole, G. L., Crawford, E. D., Tucker, M., Gerhard, D. S., Fraumeni Jr., J. F., Hoover, R., Hayes, R. B., Hunter, D. J., and Chanock, S. J. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*, 40(3):310–315.

Tol, J., Nagtegaal, I. D., and Punt, C. J. (2009). BRAF mutation in metastatic colorectal cancer. *N Engl J Med*, 361(1):98–99.

Tran, B., Kopetz, S., Tie, J., Gibbs, P., Jiang, Z. Q., Lieu, C. H., Agarwal, A., Maru, D. M., Sieber, O., and Desai, J. (2011). Impact of BRAF mutation and microsatellite instability on the pattern of metastatic spread and prognosis in metastatic colorectal cancer. *Cancer*, 117(20):4623–4632.

Traylor, M., Makela, K. M., Kilarski, L. L., Holliday, E. G., Devan, W. J., Nalls, M. A., Wiggins, K. L., Zhao, W., Cheng, Y. C., Achterberg, S., Malik, R., Sudlow, C., Bevan, S., Raitoharju, E., Oksala, N., Thijs, V., Lemmens, R., Lindgren, A., Slowik, A., Maguire, J. M., Walters, M., Algra, A., Sharma, P., Attia, J. R., Boncoraglio, G. B., Rothwell, P. M., de Bakker, P. I., Bis, J. C., Saleheen, D., Kittner, S. J., Mitchell, B. D., Rosand, J., Meschia, J. F., Levi, C., Dichgans, M., Lehtimaki, T., Lewis, C. M., and Markus, H. S. (2014). A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet*, 10(7):e1004469.

Turajlic, S., Sottoriva, A., Graham, T., and Swanton, C. (2019). Resolving genetic heterogeneity in cancer. *Nat Rev Genet*, 20(7):404–416.

Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw, A. T., de Andrade, M., Doheny, K. F., Haines, J. L., Hayes, G., Jarvik, G., Jiang, L., Kullo, I. J., Li, R., Ling, H., Manolio, T. A., Matsumoto, M., McCarty, C. A., McDavid, A. N., Mirel, D. B., Paschall, J. E., Pugh, E. W., Rasmussen, L. V., Wilke, R. A., Zuvich, R. L., and Ritchie, M. D. (2011). Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet*, Chapter 1:Unit1.19.

Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*.

Van Cutsem, E., Cervantes, A., Adam, R., Sobrero, A., Van Krieken, J. H., Aderka, D., Aranda Aguilar, E., Bardelli, A., Benson, A., Bodoky, G., Ciardiello, F., D'Hoore, A., Diaz-Rubio, E., Douillard, J. Y., Ducreux, M., Falcone, A., Grothey, A., Gruenberger, T., Haustermans, K., Heinemann, V., Hoff, P., Kohne, C. H., Labianca, R., Laurent-Puig, P., Ma, B., Maughan, T.,

Muro, K., Normanno, N., Osterlund, P., Oyen, W. J., Papamichael, D., Pentheroudakis, G., Pfeiffer, P., Price, T. J., Punt, C., Ricke, J., Roth, A., Salazar, R., Scheithauer, W., Schmoll, H. J., Tabernero, J., Taieb, J., Tejpar, S., Wasan, H., Yoshino, T., Zaanan, A., and Arnold, D. (2016). ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol*, 27(8):1386–1422.

Van Cutsem, E., Cervantes, A., Nordlinger, B., and Arnold, D. (2014). Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 25 Suppl 3:iii1–9.

Van Cutsem, E., Kohne, C. H., Lang, I., Folprecht, G., Nowacki, M. P., Cascinu, S., Shchepotin, I., Maurel, J., Cunningham, D., Tejpar, S., Schlichting, M., Zubel, A., Celik, I., Rougier, P., and Ciardiello, F. (2011). Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of overall survival according to tumor KRAS and BRAF mutation status. *J Clin Oncol*, 29(15):2011–2019.

Van Cutsem, E., Lenz, H. J., Kohne, C. H., Heinemann, V., Tejpar, S., Melezinek, I., Beier, F., Stroh, C., Rougier, P., van Krieken, J. H., and Ciardiello, F. (2015). Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. *J Clin Oncol*, 33(7):692–700.

van Heemst, D., den Reijer, P. M., and Westendorp, R. G. J. (2007). Ageing or cancer: a review on the role of caretakers and gatekeepers. *European journal of cancer (Oxford, England : 1990)*, 43(15):2144–2152.

Venderbosch, S., Nagtegaal, I. D., Maughan, T. S., Smith, C. G., Cheadle, J. P., Fisher, D., Kaplan, R., Quirke, P., Seymour, M. T., Richman, S. D., Meijer, G. A., Ylstra, B., Heideman, D. A., de Haan, A. F., Punt, C. J., and Koopman, M. (2014). Mismatch Repair Status and BRAF Mutation Status in Metastatic Colorectal Cancer Patients: A Pooled Analysis of the CAIRO, CAIRO2, COIN, and FOCUS Studies. *Clin Cancer Res*, 20(20):5322–5330.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*, 101(1):5–22.

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M., and Bos, J. L. (1988). Genetic alterations during colorectal-tumor development. *N Engl J Med*, 319(9):525–532.

Vogelstein, B. and Kinzler, K. W. (2004). Cancer genes and the pathways they control. *Nat Med*, 10(8):789–799.

Walther, A., Houlston, R., and Tomlinson, I. (2008). Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut*, 57(7):941–950.

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., and Kerr, D. (2009). Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer*, 9(7):489–499.

Wan, P. T., Garnett, M. J., Roe, S. M., Lee, S., Niculescu-Duvaz, D., Good, V. M., Jones, C. M., Marshall, C. J., Springer, C. J., Barford, D., and Marais, R. (2004). Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell*, 116(6):855–867.

Wan, S., Lai, Y., Myers, R. E., Li, B., Hyslop, T., London, J., Chatterjee, D., Palazzo, J. P., Burkart, A. L., Zhang, K., Xing, J., and Yang, H. (2013). Preoperative platelet count associates with survival and distant metastasis in surgically resected colorectal cancer patients. *J Gastrointest Cancer*, 44(3):293–304.

Wang, L., Xu, M., Qin, J., Lin, S. C., Lee, H. J., Tsai, S. Y., and Tsai, M. J. (2016). MPC1, a key gene in cancer metabolism, is regulated by COUPTFII in human prostate cancer. *Oncotarget*, 7(12):14673–14683.

Wang, W. Y., Barratt, B. J., Clayton, D. G., and Todd, J. A. (2005). Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet*, 6(2):109–118.

Wang, Y., Loree, J. M., Yu, C., Tschautscher, M., Briggler, A. M., Overman, M. J., Broaddus, R., Meric-Bernstam, F., Jones, J. C., Balcom, J., Kipp, B., Kopetz, S., and Grothey, A. (2018). Distinct impacts of KRAS, NRAS and BRAF mutations on survival of patients with metastatic colorectal cancer. *Journal of Clinical Oncology*, 36(15_suppl):3513.

Wang, Y., Velho, S., Vakiani, E., Peng, S., Bass, A. J., Chu, G. C., Gierut, J., Bugni, J. M., Der, C. J., Philips, M., Solit, D. B., and Haigis, K. M. (2013). Mutant N-RAS protects colorectal cancer cells from stress-induced apoptosis and contributes to cancer development and progression. *Cancer discovery*, 3(3):294–307.

Wasan, H., Meade, A. M., Adams, R., Wilson, R., Pugh, C., Fisher, D., Sydes, B., Madi, A., Sizer, B., Lowdell, C., Middleton, G., Butler, R., Kaplan, R., and Maughan, T. (2014). Intermittent chemotherapy plus either intermittent or continuous cetuximab for first-line treatment of patients with KRAS wild-type advanced colorectal cancer (COIN-B): a randomised phase 2 trial. *Lancet Oncol*, 15(6):631–639.

Werner, A. and Swan, D. (2010). What are natural antisense transcripts good for? *Biochem Soc Trans*, 38(4):1144–1149.

Whiffin, N., Hosking, F. J., Farrington, S. M., Palles, C., Dobbins, S. E., Zgaga, L., Lloyd, A., Kinnersley, B., Gorman, M., Tenesa, A., Broderick, P., Wang, Y., Barclay, E., Hayward, C., Martin, L., Buchanan, D. D., Win, A. K., Hopper, J., Jenkins, M., Lindor, N. M., Newcomb, P. A., Gallinger, S., Conti, D., Schumacher, F., Casey, G., Liu, T., Campbell, H., Lindblom, A., Houlston, R. S., Tomlinson, I. P., and Dunlop, M. G. (2014). Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet*, 23(17):4729–4737.

White, E., Patterson, R. E., Kristal, A. R., Thornquist, M., King, I., Shattuck, A. L., Evans, I., Satia-Abouta, J., Littman, A. J., and Potter, J. D. (2004). VITamins And Lifestyle cohort study: study design and characteristics of supplement users. *Am J Epidemiol*, 159(1):83–93.

Widmer, C., Lippert, C., Weissbrod, O., Fusi, N., Kadie, C., Davidson, R., Listgarten, J., and Heckerman, D. (2014). Further improvements to linear mixed models for genome-wide association studies. *Sci Rep*, 4:6874.

Williams, M. J., Werner, B., Heide, T., Curtis, C., Barnes, C. P., Sottoriva, A., and Graham, T. A. (2018). Quantification of subclonal selection in cancer from bulk sequencing data. *Nature genetics*, 50(6):895–903.

Wolpin, B. M. and Mayer, R. J. (2008). Systemic treatment of colorectal cancer. *Gastroenterology*, 134(5):1296–1310.

Xu, S., Kong, D., Chen, Q., Ping, Y., and Pang, D. (2017). Oncogenic long noncoding RNA landscape in breast cancer. *Mol Cancer*, 16(1):129.

Yamauchi, M., Morikawa, T., Kuchiba, A., Imamura, Y., Qian, Z. R., Nishihara, R., Liao, X., Waldron, L., Hoshida, Y., Huttenhower, C., Chan, A. T., Giovannucci, E., Fuchs, C., and Ogino, S. (2012). Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut*, 61(6):847–854.

Yang, G., Hamadeh, I. S., Katz, J., Riva, A., Lakatos, P., Balla, B., Kosa, J., Vaszilko, M., Pelliccioni, G. A., Davis, N., Langaee, T. Y., Moreb, J. S., and Gong, Y. (2018). SIRT1/HERC4 Locus Associated With Bisphosphonate-Induced Osteonecrosis of the Jaw: An Exome-Wide Association Analysis. *J Bone Miner Res*, 33(1):91–98.

Yang, Y., Zhao, L., Lei, L., Lau, W. B., Lau, B., Yang, Q., Le, X., Yang, H., Wang, C., Luo, Z., Xuan, Y., Chen, Y., Deng, X., Xu, L., Feng, M., Yi, T., Zhao, X., Wei, Y., and Zhou, S. (2017). LncRNAs: the bridge linking RNA and colorectal cancer. *Oncotarget*, 8(7):12517–12532.

Yarden, Y. (2001). The EGFR family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities. *Eur J Cancer*, 37 Suppl 4:S3–8.

Yokota, T., Ura, T., Shibata, N., Takahari, D., Shitara, K., Nomura, M., Kondo, C., Mizota, A., Utsunomiya, S., Muro, K., and Yatabe, Y. (2011). BRAF mutation is a powerful prognostic factor in advanced and recurrent colorectal cancer. *Br J Cancer*, 104(5):856–862.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A. P., and Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, 451.

Zhang, K., Civan, J., Mukherjee, S., Patel, F., and Yang, H. (2014). Genetic variations in colorectal cancer risk and clinical outcome. *World J Gastroenterol*, 20(15):4167–4177.

Zhang, L., Fan, S., Liu, H., and Huang, C. (2012). Targeting SMARCAL1 as a novel strategy for cancer therapy. *Biochem Biophys Res Commun*, 427(2):232–235.

Zheng, H. F., Rong, J. J., Liu, M., Han, F., Zhang, X. W., Richards, J. B., and Wang, L. (2015). Performance of genotype imputation for low frequency and rare variants from the 1000 genomes. *PLoS One*, 10(1):e0116487.

Zhou, W., Fritsche, L. G., Das, S., Zhang, H., Nielsen, J. B., Holmen, O. L., Chen, J., Lin, M., Elvestad, M. B., Hveem, K., Abecasis, G. R., Kang, H. M., and Willer, C. J. (2017a). Improving power of association tests using multiple sets of imputed genotypes from distributed reference panels. *Genet Epidemiol*, 41(8):744–755.

Zhou, X., Xiong, Z. J., Xiao, S. M., Zhou, J., Ding, Z., Tang, L. C., Chen, X. D., Xu, R., and Zhao, P. (2017b). Overexpression of MPC1 inhibits the proliferation, migration, invasion, and stem cell-like properties of gastric cancer cells. *Onco Targets Ther*, 10:5151–5163.

Zlobec, I., Kovac, M., Erzberger, P., Molinari, F., Bihl, M. P., Rufle, A., Foerster, A., Frattini, M., Terracciano, L., Heinimann, K., and Lugli, A. (2010). Combined analysis of specific KRAS mutation, BRAF and microsatellite instability identifies prognostic subgroups of sporadic and hereditary colorectal cancer. *Int J Cancer*, 127(11):2569–2575.

Zollner, S. and Pritchard, J. K. (2007). Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*, 80(4):605–615.

Zou, H., Chen, Q., Zhang, A., Wang, S., Wu, H., Yuan, Y., Wang, S., Yu, J., Luo, M., Wen, X., Cui, W., Fu, W., Yu, R., Chen, L., Zhang, M., Lan, H., Zhang, X., Xie, Q., Jin, G., and Xu, C. (2019). MPC1 deficiency accelerates lung adenocarcinoma progression through the STAT3 pathway. *Cell Death Dis*, 10(3):148.

Appendices

Published OnlineFirst November 4, 2016; DOI: 10.1158/1078-0432.CCR-16-1541

Personalized Medicine and Imaging

Clinical Cancer Research

BRAF and *NRAS* Locus-Specific Variants Have Different Outcomes on Survival to Colorectal Cancer

Matthew G. Summers[1], Christopher G. Smith[1], Timothy S. Maughan[2], Richard Kaplan[3], Valentina Escott-Price[4], and Jeremy P. Cheadle[1]

Abstract

Purpose: Somatic mutation status at *KRAS, BRAF,* and *NRAS* is associated with prognosis in patients with advanced colorectal cancer (aCRC); however, it remains unclear whether there are intralocus, variant-specific differences in survival and other clinicopathologic parameters.

Experimental Design: We profiled 2,157 aCRCs for somatic mutations in *KRAS, BRAF,* and *NRAS* and determined microsatellite instability status. We sought inter- and intralocus correlations between mutations and variant-specific associations with survival and clinicopathology.

Results: *KRAS* mutations were rarely found together and those in codons 12 and 13 conferred poor prognosis [hazard ratio (HR), 1.44; 95% confidence interval (CI), 1.28–1.61; $P = 6.4 \times 10^{-10}$ and HR, 1.53; 95% CI, 1.26–1.86; $P = 1.5 \times 10^{-05}$, respectively]. For *BRAF,* more c.1781A>G (p.D594G) CRCs carried *RAS* mutations [14% (3/21)] compared with c.1799T>A (p.V600E) CRCs [1% (2/178), $P = 9.0 \times 10^{-03}$]. c.1799T>A (p.V600E) was associated with poor prognosis (HR, 2.60; 95% CI, 2.06–3.28; $P = 1.0 \times 10^{-15}$), whereas c.1781A>G (p.D594G) was not (HR, 1.30; 95% CI, 0.73–2.31; $P = 0.37$); this intralocus difference was significant ($P = 0.04$). More c.1799T>A (p.V600E) colorectal cancers were found in the right colon [47% (47/100)], compared with c.1781A>G (p.D594G) colorectal cancers [7% (1/15), $P = 3.7 \times 10^{-03}$]. For *NRAS,* 5% (3/60) of codon 61 mutant colorectal cancers had *KRAS* mutations compared with 44% (10/23) of codons 12 and 13 mutant colorectal cancers ($P = 7.9 \times 10^{-05}$). Codon 61 mutations conferred poor prognosis (HR, 1.47; 95% CI, 1.09–1.99; $P = 0.01$), whereas codons 12 and 13 mutations did not (HR, 1.29; 95% CI, 0.64–2.58; $P = 0.48$).

Conclusions: Our data show considerable intralocus variation in the outcomes of mutations in *BRAF* and *NRAS*. These data need to be considered in patient management and personalized cancer therapy. *Clin Cancer Res; 23(11); 2742–9. ©2016 AACR.*

Introduction

The only routinely used prognostic marker for survival after diagnosis of colorectal cancer is clinical stage, which combines depth of tumor invasion, nodal status, and distant metastasis (1). In stage IV disease, Köhne's index based on performance status, white blood cell count, alkaline phosphatase levels, and number of metastatic sites has been proposed (2). Other factors thought to influence survival include lifestyle (3, 4), systemic inflammatory response to the tumor (5), tumor immunologic environment (6), and the germline (7) and somatic (8–11) molecular profiles. By studying patients with advanced colorectal cancer (aCRC) from the Medical Research Council (MRC) COIN trial, we previously showed that the somatic mutation status at *KRAS* and *BRAF*, and microsatellite instability (MSI), conferred poor prognosis irrespective of treatment: overall survival (OS, trial enrolment to death) *KRAS*-mutant 14.4 months (12), *BRAF*-mutant 8.8 months (12), MSI 9.3 months (13), all wild-type 20.1 months (12). We also showed that neither individual somatic mutations nor mutations grouped by codon or gene affected response to cetuximab (13).

It remains unclear whether there are intralocus, variant-specific differences in survival, and this has been difficult to study for the less frequently mutated loci [such as c.1781A>G (p.D594G) in *BRAF*] due to the large numbers of samples required to make statistically robust associations. Here, we studied the influence of individual or codon-specific somatic mutations in *KRAS, BRAF,* and *NRAS* in 2,157 patients with aCRC from COIN (12) and COIN-B (14).

Materials and Methods

Patients and samples

We prepared tumor DNA samples from unrelated patients with aCRC from the MRC clinical trials COIN (NCT00182715; ref. 12) and COIN-B (NCT00640081; ref. 14), as described previously

[1]Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff, United Kingdom. [2]CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Oxford, United Kingdom. [3]MRC Clinical Trials Unit, London, United Kingdom. [4]Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, United Kingdom.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

V. Escott-Price and J.P. Cheadle share senior authorship of this article.

Current address for C.G. Smith: CRUK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, United Kingdom.

Corresponding Author: Jeremy P. Cheadle, Division of Cancer and Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, United Kingdom. Phone: 4429-2074-2652; E-mail: cheadlejp@cardiff.ac.uk

doi: 10.1158/1078-0432.CCR-16-1541

Published OnlineFirst November 4, 2016; DOI: 10.1158/1078-0432.CCR-16-1541

Translational Relevance

Somatic mutation status at *KRAS*, *BRAF*, and *NRAS* affects prognosis in patients with advanced colorectal cancer (aCRC), and it has been presumed that different variants in the same gene confer similar prognostic outcomes. Here, we studied inter- and intralocus variant cooccurrence and variant-specific differences in survival and clinicopathology by analyzing 2,157 patients with aCRC. We found significant differences between variants in *BRAF* [c.1781A>G (p.D594G) versus c.1799T>A (p.V600E)] and *NRAS* (mutant codons 12 and 13 vs. codon 61) both in terms of cooccurrence with *KRAS* mutations and in their influence on survival. These data need to be considered in patient management and personalized therapy.

(12, 13). All patients had either previous or current histologically confirmed primary adenocarcinomas of the 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 tumor. COIN patients were randomized 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy, continuous chemotherapy plus cetuximab, or intermittent chemotherapy. COIN-B patients were randomized 1:1 to receive intermittent chemotherapy plus continuous cetuximab or intermittent chemotherapy plus intermittent cetuximab. All patients gave informed consent for their samples to be used for bowel cancer research [approved by REC (04/MRE06/60)].

Somatic analyses

We previously screened for somatic mutations in *KRAS* (codons 12, 13, and 61), *BRAF* (codons 594 and 600), and *NRAS* (codons 12, 13, and 61) using a combination of pyrosequencing and Sequenom (13); for samples analyzed by both technologies (n = 1,612), genotype concordance in *KRAS* was 99% (8,642/8,719 calls were concordant). MSI status was determined using the markers BAT-25 and BAT-26 (13).

Mutation cooccurrence, survival, and statistical analyses

We sought inter- and intralocus correlations between somatic *KRAS*, *BRAF*, and *NRAS* mutations and MSI status. Data were analyzed using R (http://www.r-project.org). Corrplot was used to create a correlation matrix plot (recode from car was used to recode the data into binary format) and Surv, survfit, survdiff, and coxph from the survival package and ggsurv from the GGally package were used to create and analyze the survival curves. To avoid potential confounding effects from other mutant loci, *KRAS* mutants (vs. wild-type) were analyzed on a *BRAF* and *NRAS*

Table 1. Prognostic outcomes of individual mutations, or mutations grouped by codon or gene on OS

Gene/event	Mutation/codon	No of events[a]	HR	95% CIs	*P*
KRAS	c.34G>A (p.G12S)	35	1.78	1.27–2.50	9.2×10^{-04} (0.03)
	c.34G>C (p.G12R)	10	0.95	0.51–1.78	0.88
	c.34G>T (p.G12C)	52	1.21	0.91–1.60	0.18
	c.35G>A (p.G12D)	187	1.48	1.26–1.74	1.6×10^{-06} (4.8×10^{-05})
	c.35G>C (p.G12A)	41	1.43	1.04–1.96	0.03
	c.35G>T (p.G12V)	161	1.48	1.25–1.76	7.5×10^{-06} (2.3×10^{-04})
	c.37G>T (p.G13C)	6	1.36	0.61–3.03	0.46
	c.38G>A (p.G13D)	116	1.53	1.26–1.87	2.2×10^{-05} (6.6×10^{-04})
	c.38G>T (p.G13V)	1	—	—	—
	c.182A>G (p.Q61R)	6	1.41	0.63–3.15	0.41
	c.182A>T (p.Q61L)	6	1.27	0.57–2.84	0.56
	c.183A>C (p.Q61H)	15	1.17	0.70–1.95	0.56
	Codon 12	486	1.44	1.28–1.61	6.4×10^{-10} (1.9×10^{-08})
	Codon 13	123	1.53	1.26–1.86	1.5×10^{-05} (4.5×10^{-04})
	Codon 61	27	1.23	0.84–1.81	0.28
	Any *KRAS* mutation	632	1.45	1.30–1.61	1.9×10^{-11} (5.7×10^{-10})
BRAF	c.1781A>G (p.D594G)	12	1.30	0.73–2.31	0.37
	c.1799T>A (p.V600E)	87	2.60	2.06–3.28	1.0×10^{-15} (3.0×10^{-14})
	Any *BRAF* mutation	99	2.31	1.85–2.87	7.8×10^{-14} (2.3×10^{-12})
NRAS	c.34G>T (p.G12C)	5	1.42	0.59–3.43	0.43
	c.35G>A (p.G12D)	2	—	—	—
	c.35G>T (p.G12V)	1	—	—	—
	c.181C>A (p.Q61K)	21	1.43	0.96–2.21	0.11
	c.182A>G (p.Q61R)	13	1.58	0.91–2.73	0.11
	c.182A>T (p.Q61L)	11	1.51	0.83–2.73	0.18
	Codons 12 and 13	8	1.29	0.64–2.58	0.48
	Codon 61	45	1.47	1.09–1.99	0.01
	Any *NRAS* mutation	53	1.44	1.09–1.90	0.01
MSI	MSI	23	1.86	1.22–2.83	4.0×10^{-03}
	MSS	476	1.00	Ref.	Ref.

NOTE: *KRAS* mutants (vs. wild-type) were analyzed on a *BRAF* and *NRAS* wild-type background; *BRAF* mutants (vs. wild-type) were analyzed on a *RAS* wild-type and MSS background; *NRAS* mutants (vs. wild-type) were analyzed on a *KRAS* and *BRAF* wild-type background; and MSI (vs. MSS) was analyzed on a *RAS* and *BRAF* wild-type background.

Number of events, HR, CIs, and *P* values are shown (except for cases where number of events ≤2).

[a]Mutations not listed when number of events = 0.

P values that remained significant after correction for multiple testing are shown in parentheses.

Published OnlineFirst November 4, 2016; DOI: 10.1158/1078-0432.CCR-16-1541

wild-type background; *BRAF* mutants (vs. wild-type) were analyzed on a *KRAS* and *NRAS* (*RAS*) wild-type and microsatellite stable (MSS) background; *NRAS* mutants (vs. wild-type) were analyzed on a *KRAS* and *BRAF* wild-type background; and MSI (vs. MSS) was analyzed on a *RAS* and *BRAF* wild-type background. We found no evidence of heterogeneity in OS between patients when analyzed by trial (COIN vs. COIN-B, $P = 0.49$), trial arm ($P = 0.40$; Cochran Q test: $P = 1.0$, I^2 test: $P = 0.74$), type of chemotherapy received (OxMdG/XELOX; $P = 0.60$), or cetuximab use ($P = 0.41$), so we combined these groups for the survival

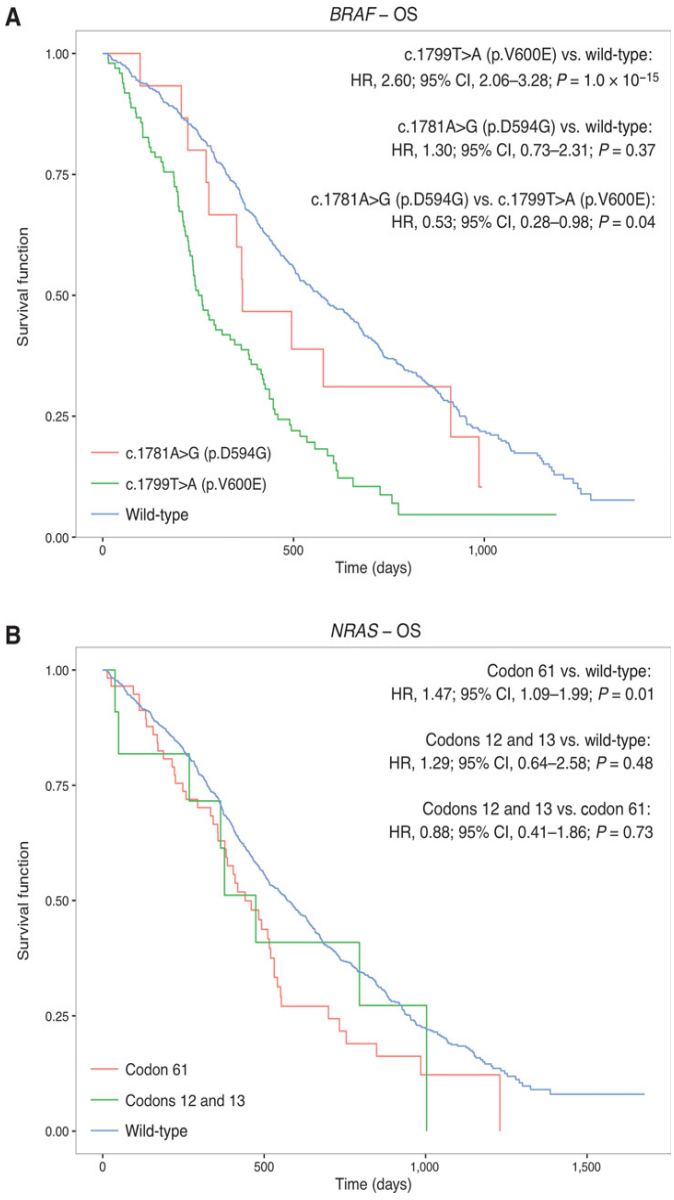

Figure 1.
Kaplan–Meier plots showing the prognostic outcome of c.1781A>G (p.D594G) and c.1799T>A (p.V600E) in *BRAF* (**A**), and codons 12 and 13 and codon 61 mutations in *NRAS* (**B**).

Published OnlineFirst November 4, 2016; DOI: 10.1158/1078-0432.CCR-16-1541

analyses. We used χ^2 tests or Fisher exact test to study whether *KRAS*, *BRAF*, and *NRAS* mutations and MSI status were associated with different clinicopathologic findings. We corrected for multiple testing using Bonferroni correction [$P < 1.7 \times 10^{-03}$ ($n = 30$) for survival tests, $P < 1.0 \times 10^{-04}$ ($n = 480$) for somatic mutation cross-correlations, $P < 1.3 \times 10^{-03}$ ($n = 39$) for clinicopathologic analyses of *KRAS*, *BRAF*, and *NRAS* and $P < 3.8 \times 10^{-03}$ ($n = 13$) for clinicopathologic analyses of MSI].

Results

We screened for somatic *KRAS*, *BRAF*, and *NRAS* mutations and for MSI status in aCRCs from 2,157 patients from the clinical trials COIN and COIN-B. In total, we detected 14 *KRAS* mutations [c.34G>A (p.G12S), c.34G>C (p.G12R), c.34G>T (p.G12C), c.35G>A (p.G12D), c.35G>C (p.G12A), c.35G>T (p.G12V), c.37G>A (p.G13S), c.37G>C (p.G13R), c.37G>T (p.G13C), c.38G>A (p.G13D), c.38G>T (p.G13V), c.182A>G (p.Q61R), c.182A>T (p.Q61L), and c.183A>C (p.Q61H)] in 40% (858/2,157) aCRCs, 2 *BRAF* mutations [c.1781A>G (p.D594G) and c.1799T>A (p.V600E)] in 9% (199/2,097), 9 *NRAS* mutations [c.34G>T (p.G12C), c.35G>A (p.G12D), c.35G>T (p.G12V), c.37G>C (p.G13R), c.38G>A (p.G13D), c.181C>A (p.Q61K), c.182A>G (p.Q61R), c.182A>T (p.Q61L), and c.183A>C (p.Q61H)] in 4% (83/2,092), and MSI in 4% (66/1,567). Over 99% (2,152/2,157) of aCRCs harboring *KRAS*, *BRAF*, and *NRAS* mutations carried only a single variant allele at their respective loci (five colorectal cancers carried two *KRAS* mutations; however, due to their rarity, these were likely to reflect mixed tumor populations).

Inter- and intragenic mutation correlations

All mutations in *KRAS*, regardless of whether analyzed individually or by codon, showed similar effects in terms of mutual exclusivity (Supplementary Fig. S1). Codon 12 (4/627 mutant colorectal cancers), 13 (4/161), and 61 (2/35) mutations were rarely found together.

Only specific mutations in *BRAF* [c.1799T>A (p.V600E)] and *NRAS* (codon 61 mutations) shared this characteristic. Only 1% (2/178) of *BRAF* c.1799T>A (p.V600E) colorectal cancers had *RAS* mutations compared with 47% (894/1,908) of *BRAF* wild-type colorectal cancers ($P < 2.2 \times 10^{-16}$, $P < 1.1 \times 10^{-13}$ after correction for multiple testing). In contrast, more *BRAF* c.1781A>G (p.D594G) mutations cooccurred with *RAS* mutations [14% (3/21)] as compared with c.1799T>A (p.V600E; $P = 9.0 \times 10^{-03}$), albeit less commonly than found in *BRAF* wild-type colorectal cancers ($P = 3.0 \times 10^{-03}$). We noted one case of *KRAS* c.37G>A (p.G13S), which cooccurred with *BRAF* c.1799T>A [p.V600E; $P = 2.5 \times 10^{-03}$ as compared with other *KRAS* mutations (1/812 cooccurred)]. For *NRAS*, only 5% (3/60) of codon 61 mutant colorectal cancers had *KRAS* mutations compared with 43% (10/23) of codons 12 and 13 mutant colorectal cancers ($P = 7.9 \times 10^{-05}$, $P = 0.04$ after correction), the latter being at a similar level to that found in wild-type colorectal cancers [40% (808/2,018), $P = 0.98$].

We also observed differences in the relationship between *BRAF* mutations and MSI status. *BRAF* c.1799T>A (p.V600E) was strongly associated with MSI [11% (20/178) c.1799T>A (p.V600E) colorectal cancers had MSI compared with 2% (46/1,908) wild-type colorectal cancers, $P = 5.3 \times 10^{-10}$, $P = 2.5 \times 10^{-07}$ after correction], whereas *BRAF* c.1781A>G (p.D594G) and MSI did not cooccur (0/21).

Table 2. Clinicopathology according to *KRAS* mutation status

Characteristics		Frequency of *KRAS* mutations[a]	Codons 12 and 13 ($n = 760$)	Codon 61 ($n = 33$)	Wild-type ($n = 1,002$)	*P* (codons 12 and 13 vs. wild-type)	*P* (codon 61 vs. wild-type)	*P* (codons 12 and 13 vs. codon 61)
Sex	Female	289/593 (49)	275 (36)	14 (42)	304 (30)	0.01	0.20	0.59
	Male	503/1201 (42)	485 (64)	19 (58)	698 (70)	0.01	0.20	0.59
Age	Mean	NA	63	61	63	NA	NA	NA
Primary tumor site	Right colon	182/314 (58)	173 (23)	9 (27)	132 (13)	1.9×10^{-07} [7.4×10^{-06}]	0.04	0.70
	Cecum	62/88 (70)	61 (8)	1 (3)	26 (3)	3.4×10^{-07} [1.3×10^{-05}]	0.59	0.51
	Transverse colon	14/35 (40)	14 (2)	0 (0)	21 (2)	0.84	1.0	1.0
	Left colon	123/326 (38)	117 (15)	6 (18)	203 (20)	0.01	0.94	0.85
	Sigmoid colon	44/159 (28)	41 (5)	3 (9)	115 (11)	1.3×10^{-05} [5.1×10^{-04}]	1.0	0.42
	Rectosigmoid junction	108/269 (40)	105 (14)	3 (9)	161 (16)	0.22	0.34	0.61
	Rectum	251/577 (44)	241 (32)	10 (30)	326 (33)	0.75	0.94	1.0
Site of metastases[b]	Liver only	156/418 (37)	152 (20)	4 (12)	262 (26)	3.1×10^{-03}	0.07	0.37
	Liver	598/1,356 (44)	577 (76)	22 (67)	758 (76)	0.94	0.33	0.32
	Nodal	359/832 (43)	345 (45)	15 (45)	473 (47)	0.48	0.98	1.0
	Lung	358/715 (50)	342 (45)	16 (48)	357 (36)	8.4×10^{-05} [3.3×10^{-03}]	0.18	0.83
	Peritoneum	126/259 (49)	117 (15)	9 (27)	133 (13)	0.23	0.04	0.11

NOTE: Mutations were analyzed on an *NRAS* and *BRAF* wild-type background.
Abbreviation: NA, not applicable.
[a]There was a significant difference between *KRAS*-mutant colorectal cancers in the location of the primary tumor ($P = 6.4 \times 10^{-14}$) and in the sites of metastases ($P = 4.6 \times 10^{-04}$), as compared with wild-type colorectal cancers.
[b]Some patients had multiple metastases, so percentages do not add up to 100%.
Percentages are shown in regular parentheses.
P values that remained significant after correction for multiple testing are shown in square parentheses.
Discrepancies in column totals are due to patients with multiple mutations or due to missing data.

Published OnlineFirst November 4, 2016; DOI: 10.1158/1078-0432.CCR-16-1541

Survival analyses

Five *KRAS* mutations [c.34G>A (p.G12S), c.35G>A (p.G12D), c.35G>C (p.G12A), c.35G>T (p.G12V), and c.38G>A (p.G13D)] individually showed significantly poorer prognosis with a median reduction in survival of 213, 111, 65, 160, and 165 days, respectively; four of these remained significant after correction for multiple testing (Table 1). When grouped by codons, both codon 12 and 13 mutations conferred poor prognosis [hazard ratio (HR), 1.44; 95% confidence interval (CI), 1.28–1.61; $P = 6.4 \times 10^{-10}$, $P = 1.9 \times 10^{-08}$ after correction, and HR, 1.53; 95% CI, 1.26–1.86; $P = 1.5 \times 10^{-05}$, $P = 4.5 \times 10^{-04}$ after correction, respectively], whereas codon 61 mutations did not (HR, 1.23; 95% CI, 0.84–1.81; $P = 0.28$; Table 1); these intralocus differences were not significant.

c.1799T>A (p.V600E) in *BRAF* was strongly associated with poor prognosis (HR, 2.60; 95% CI, 2.06–3.28; $P = 1.0 \times 10^{-15}$, $P = 3.0 \times 10^{-14}$ after correction, median reduction in survival 320 days; Fig. 1), whereas c.1781A>G (p.D594G) was not (HR, 1.30; 95% CI, 0.73–2.31; $P = 0.37$); this intralocus difference was significant ($P = 0.04$; Table 1).

Although individual *NRAS* mutations showed no differences in survival, when grouped by codon, codon 61 mutations conferred poor prognosis (HR, 1.47; 95% CI, 1.09–1.99; $P = 0.01$, median reduction in survival 131 days; Fig. 1), whereas codons 12 and 13 mutations did not (HR, 1.29; 95% CI, 0.64–2.58; $P = 0.48$); however, this intralocus difference was not significant ($P = 0.73$).

Patients with MSI colorectal cancers had worse prognosis compared with those with stable tumors (HR, 1.86; 95% CI, 1.22–2.83; $P = 4.0 \times 10^{-03}$), in agreement with our previous study (13).

For all analyses described herein, there were no significant differences measured using heterogeneity tests when the analyses were performed using date of diagnosis to death instead of OS (Supplementary Table S1) or when split by cetuximab use (Supplementary Table S2).

Clinicopathologic analyses

KRAS. More *KRAS*-mutant colorectal cancers were found in the right colon [58% (182/314)] and cecum [70% (62/88)] as compared with the left colon [38% (123/326), $P = 4.6 \times 10^{-07}$ and 8.4×10^{-08}, respectively], and more were associated with metastases in the lung [50% (358/715)] as compared with the liver only [37% (156/418), $P = 4.2 \times 10^{-05}$; Table 2].

In terms of codon-specific mutations, more *KRAS* codon 12 and 13 mutant colorectal cancers were found in the right colon [23% (173/760) vs. 13% (132/1,002), $P = 1.9 \times 10^{-07}$] and cecum [8% (61/760) vs. 3% (26/1,002), $P = 3.4 \times 10^{-07}$], less in the left colon [15% (117/760) vs. 20% (203/1,002), $P = 0.01$] and sigmoid colon [5% (41/760) vs. 11% (115/1,002), $P = 1.3 \times 10^{-05}$], and more were associated with metastases in the lung [45% (342/760) vs. 36% (357/1,002), $P = 8.4 \times 10^{-05}$] and less in liver only [20% (152/760) vs. 26% (262/1,002), $P = 3.1 \times 10^{-03}$], as compared with wild-type colorectal cancers; the correlations for right colon, cecum, sigmoid colon, and lung remained significant after correction for multiple testing (Table 2). More *KRAS* codon 61 mutant patients had colorectal cancers in the right colon [27% (9/33) vs. 13% (132/1,002), $P = 0.04$] and more had peritoneal metastases [27% (9/33) vs. 13% (133/1,002), $P = 0.04$] as compared with wild-type patients. However, there were no significant differences in clinicopathology between *KRAS* codons 12 and 13 versus codon 61 mutant patients.

Table 3. Clinicopathology according to *BRAF* mutation status

Characteristics		Frequency of *BRAF* mutations[a]	c.1781A>G (p.D594G; $n = 15$)	c.1799T>A (p.V600E; $n = 100$)	Wild-type ($n = 693$)	*P* (c.1781A>G [p.D594G] vs. wild-type)	p (c.1799T>A [p.V600E] vs. wild-type)	p (c.1781A>G [p.D594G] vs. c.1799T>A [p.V600E])
Sex	Female	55/249 (22)	7 (47)	48 (48)	194 (28)	0.20	8.0×10^{-05} [3.1×10^{-03}]	1.0
	Male	60/559 (11)	8 (53)	52 (52)	499 (72)	0.20	8.0×10^{-05} [3.1×10^{-03}]	1.0
Age	Mean	NA	67	63	63	NA	NA	NA
Primary tumor site	Right colon	48/128 (38)	1 (7)	47 (47)	80 (12)	1.0	$<2.2 \times 10^{-16}$ [$<8.6 \times 10^{-15}$]	3.7×10^{-03}
	Cecum	4/24 (17)	0 (0)	4 (4)	20 (3)	1.0	0.53	1.0
	Transverse colon	4/20 (20)	0 (0)	4 (4)	16 (2)	1.0	0.30	1.0
	Left colon	18/146 (12)	1 (7)	17 (17)	128 (18)	0.34	0.83	0.46
	Sigmoid colon	4/77 (5)	2 (13)	2 (2)	73 (11)	0.67	3.2×10^{-03}	0.08
	Rectosigmoid junction	15/141 (11)	2 (13)	13 (13)	126 (18)	1.0	0.26	1.0
	Rectum	20/254 (8)	9 (60)	11 (11)	234 (34)	0.07	7.1×10^{-06} [2.8×10^{-04}]	1.7×10^{-05} [6.6×10^{-04}]
Site of metastases[b]	Liver only	22/214 (10)	3 (20)	19 (19)	192 (28)	0.77	0.09	1.0
	Liver	83/619 (13)	13 (87)	70 (70)	536 (77)	0.54	0.14	0.23
	Nodal	53/368 (14)	7 (47)	46 (46)	315 (45)	1.0	1.0	1.0
	Lung	35/272 (13)	6 (40)	29 (29)	237 (34)	0.85	0.36	0.57
	Peritoneum	25/107 (23)	1 (7)	24 (24)	82 (12)	1.0	1.5×10^{-03}	0.19

NOTE: Mutations analyzed on a *RAS* wild-type and MSS background.
Abbreviation: NA, not applicable.
[a]There was a significant difference between *BRAF*-mutant colorectal cancers in the location of the primary tumor ($P = 1.2 \times 10^{-13}$) and in the sites of metastases ($P = 0.03$) as compared with wild-type colorectal cancers.
[b]Some patients had multiple metastases so percentages do not add up to 100%.
Percentages are shown in regular parentheses.
P values that remained significant after correction for multiple testing are shown in square parentheses.
Discrepancies in column totals are due to patients with multiple mutations or due to missing data.

Published OnlineFirst November 4, 2016; DOI: 10.1158/1078-0432.CCR-16-1541

Table 4. Clinicopathology according to *NRAS* mutation status

Characteristics		Frequency of *NRAS* mutations[a]	Codons 12 and 13 ($n = 11$)	Codon 61 ($n = 57$)	Wild-type ($n = 1{,}002$)	*P* (codons 12 and 13 vs. wild-type)	*P* (codon 61 vs. wild-type)	*P* (codons 12 and 13 vs. codon 61)
Sex	Female	20/324 (6)	2 (18)	18 (32)	304 (30)	0.52	0.96	0.49
	Male	48/746 (6)	9 (82)	39 (68)	698 (70)	0.52	0.96	0.49
Age	Mean	NA	59	62	63	NA	NA	NA
Primary tumor site	Right colon	5/137 (4)	2 (18)	3 (5)	132 (13)	0.65	0.10	0.18
	Cecum	4/30 (13)	0 (0)	4 (7)	26 (3)	1.0	0.07	1.0
	Transverse colon	3/24 (13)	0 (0)	3 (5)	21 (2)	1.0	0.13	1.0
	Left colon	12/215 (6)	1 (9)	11 (19)	203 (20)	0.70	1.0	0.67
	Sigmoid colon	11/126 (9)	2 (18)	9 (16)	115 (11)	0.37	0.44	1.0
	Rectosigmoid junction	10/171 (6)	0 (0)	10 (18)	161 (16)	0.23	0.91	0.20
	Rectum	19/345 (6)	6 (55)	13 (23)	326 (33)	0.22	0.17	0.08
Site of metastases[b]	Liver only	10/272 (4)	3 (27)	7 (12)	262 (26)	1.0	0.03	0.35
	Liver	52/810 (6)	8 (73)	44 (77)	758 (76)	0.74	0.92	0.71
	Nodal	35/508 (7)	3 (27)	32 (56)	473 (47)	0.23	0.24	0.11
	Lung	43/400 (11)	4 (36)	39 (68)	357 (36)	1.0	1.3×10^{-06} [5.1×10^{-05}]	0.08
	Peritoneum	5/138 (4)	1 (9)	4 (7)	133 (13)	1.0	0.22	1.0

NOTE: Mutations analyzed on a *KRAS* and *BRAF* wild-type background.
Abbreviation: NA, not applicable.
[a]There was a significant difference between *NRAS*-mutant colorectal cancers in the sites of metastases ($P = 2.5 \times 10^{-03}$) as compared with wild-type colorectal cancers.
[b]Some patients had multiple metastases so percentages do not add up to 100%.
Percentages are shown in regular parentheses.
P values that remained significant after correction for multiple testing are shown in square parentheses.
Discrepancies in column totals are due to patients with multiple mutations or due to missing data.

Table 5. Clinicopathology according to MSI status

Characteristics		Frequency of MSI[a]	MSI ($n = 29$)	MSS ($n = 693$)	*P* (MSI vs. MSS)
Sex	Female	11/205 (5)	11 (38)	194 (28)	0.34
	Male	18/517 (3)	18 (62)	499 (72)	0.34
Age	Mean	NA	58	63	NA
Primary tumor site	Right colon	12/92 (13)	12 (41)	80 (12)	9.2×10^{-06} [1.2×10^{-04}]
	Cecum	1/21 (5)	1 (3)	20 (3)	0.58
	Transverse colon	1/17 (6)	1 (3)	16 (2)	0.51
	Left colon	7/135 (5)	7 (24)	128 (18)	0.60
	Sigmoid colon	1/76 (1)	1 (3)	73 (11)	0.35
	Rectosigmoid junction	1/127 (1)	1 (3)	126 (18)	0.04
	Rectum	6/240 (3)	6 (21)	234 (34)	0.21
Site of metastases[b]	Liver only	3/195 (2)	3 (10)	192 (28)	0.05
	Liver	14/550 (3)	14 (48)	536 (77)	7.3×10^{-04} [9.5×10^{-03}]
	Nodal	17/332 (5)	17 (59)	315 (45)	0.23
	Lung	6/243 (2)	6 (21)	237 (34)	0.19
	Peritoneum	7/89 (8)	7 (24)	82 (12)	0.09

NOTE: MSI status was analyzed on an *RAS* and *BRAF* wild-type background.
Abbreviation: NA, not applicable.
[a]There was a significant difference between MSI colorectal cancers in the location of the primary tumor ($P = 2.5 \times 10^{-04}$) and in the sites of metastases ($P = 0.02$) as compared with MSS colorectal cancers.
[b]Some patients had multiple metastases so percentages do not add up to 100%.
Percentages are shown in regular parentheses.
P values that remained significant after correction for multiple testing are shown in square parentheses.
Discrepancies in column totals are due to patients with multiple mutations or due to missing data.

Published OnlineFirst November 4, 2016; DOI: 10.1158/1078-0432.CCR-16-1541

BRAF. More *BRAF*-mutant colorectal cancers were found in the right colon [38% (48/128)] as compared with the left colon [12% (18/146), $P = 2.4 \times 10^{-05}$], and more were associated with metastases in the peritoneum [23% (25/107)] as compared with liver only [10% (22/214), $P = 3.1 \times 10^{-03}$; Table 3].

In terms of individual mutations, *BRAF* c.1781A>G (p.D594G) colorectal cancers had similar clinicopathology to wild-type colorectal cancers (Table 3). In contrast, more *BRAF* c.1799T>A (p.V600E) colorectal cancers were found in the right colon [47% (47/100) vs. 12% (80/693), $P < 2.2 \times 10^{-16}$], and less in the rectum [11% (11/100) vs. 34% (234/693), $P = 7.1 \times 10^{-06}$] and sigmoid colon [2% (2/100) vs. 11% (73/693), $P = 3.2 \times 10^{-03}$], and more were associated with peritoneal metastases [24% (24/100) vs. 12% (82/693), $P = 1.5 \times 10^{-03}$] as compared with wild-type colorectal cancers; the correlations for right colon and rectum remained significant after correction for multiple testing (Table 3).

In terms of intralocus differences, there was a significant difference between c.1781A>G (p.D594G) and c.1799T>A (p.V600E) colorectal cancers in the location of the primary tumor ($P = 9.3 \times 10^{-05}$, $P = 3.6 \times 10^{-03}$ after correction), due to fewer c.1781A>G (p.D594G) colorectal cancers in the right colon [7% (1/15) vs. 47% (47/100), $P = 3.7 \times 10^{-03}$] and more in the rectum [60% (9/15) vs. 11% (11/100), $P = 1.7 \times 10^{-05}$, $P = 6.6 \times 10^{-04}$ after correction; Table 3]. There was no significant difference between the sites of metastases associated with these mutations.

NRAS. There was no difference between the frequency of *NRAS*-mutant and wild-type colorectal cancers in the site of the primary tumor (Table 4). However, more *NRAS*-mutant colorectal cancers were associated with metastases in the lung [11% (43/400)] as compared with the liver only [4% (10/272), $P = 1.4 \times 10^{-03}$].

In terms of individual codons, codon 12 and 13 mutant colorectal cancers showed similar clinicopathology to wild-type colorectal cancers (Table 4). Codon 61 mutant colorectal cancers had similar primary tumor distributions but significantly fewer liver only [12% (7/57) vs. 26% (262/1,002), $P = 0.03$] and more lung metastases [68% (39/57) vs. 36% (357/1,002), $P = 1.3 \times 10^{-06}$, $P = 5.1 \times 10^{-05}$ after correction] as compared with wild-type colorectal cancers (Table 4). There were no significant differences in clinicopathology between codons 12 and 13 vs. codon 61 mutant colorectal cancers.

MSI. More MSI colorectal cancers were found in the right colon [41% (12/29) vs. 12% (80/693), $P = 9.2 \times 10^{-06}$, $P = 1.2 \times 10^{-04}$ after correction] and less in the rectosigmoid junction [3% (1/29) vs. 18% (126/693), $P = 0.04$], and less were associated with liver metastases [48% (14/29) vs. 77% (536/693), $P = 7.3 \times 10^{-04}$, $P = 9.5 \times 10^{-03}$ after correction] as compared with MSS colorectal cancers (Table 5).

Discussion

Variants in *BRAF* and *NRAS* have been presumed to confer similar oncogenic and prognostic outcomes; however, here we demonstrate clear intralocus differences. For *BRAF*, c.1799T>A (p.V600E) was almost mutually exclusive of *RAS* mutations and was associated with poor prognosis. In contrast, c.1781A>G (p.D594G) was more often associated with *RAS* mutations and had no apparent influence on survival. However, c.1781A>G (p.D594G) is unlikely to be benign and more likely to be hypomorphic, as it had significantly fewer cooccurrences with *RAS* mutations as compared with *BRAF* wild-type colorectal cancers. Interestingly, our data are consistent with a recent report showing that patients with codon 594 or 596 mutated tumors had longer OS compared with those with c.1799T>A (p.V600E) colorectal cancers (15). There are clear biological differences between these mutant codons to support our observed pathologic differences; p.V600E increased ERK and NFκB signaling and the transformation of NIH3T3 cells, whereas p.D594V failed to activate ERK (16) and did not affect NFκB signaling nor NIH3T3 transforming activity (17).

Others have reported that *NRAS*-mutant patients have shorter OS as compared with wild-type patients (HR, 1.91; 95% CI, 1.39–3.86; $P = 1.3 \times 10^{-03}$; ref. 18). Here, we noted a more complex relationship; *NRAS* codon 61 mutations, which were rarely associated with *KRAS* mutations, conferred a poor prognosis, but codons 12 and 13 mutations, which cooccurred with *KRAS* mutations at similar frequencies to wild-type colorectal cancers, had little influence on survival. Together, our data suggest that *NRAS* codons 12 and 13 mutations may have a minor role in colorectal tumorigenesis. Interestingly, using mouse models, others have shown that endogenous levels of *Nras* p.Q61R, but not *Nras* p.G12D, were able to efficiently drive *in vivo* melanomagenesis (19), supporting their differing biological effects.

We have also shown that different mutant loci are associated with differences in the clinicopathology of the primary tumors and/or their sites of metastases. For example, in agreement with two recent reports (20, 21), we observed more *KRAS*-mutant colorectal cancers in the cecum (70%) and, to a lesser extent, in the right colon (58%), as compared with the left colon (38%). It has been suggested that different somatic profiles are associated with different clinicopathology, by influencing the tumor's biological behavior (22). Here, we focused on intralocus differences and found a significant difference between c.1781A>G (p.D594G) and c.1799T>A (p.V600E) in *BRAF* in the location of the primary tumor providing additional support for these variants having different biological effects.

In conclusion, our study shows considerable intralocus variations in survival, particularly in the outcomes of mutations in *BRAF* and *NRAS*. These data need to be considered in patient management.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: M.G. Summers, T.S. Maughan, V. Escott-Price, J.P. Cheadle

Development of methodology: M.G. Summers, J.P. Cheadle

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.G. Summers, C.G. Smith, T.S. Maughan, R. Kaplan, J.P. Cheadle

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M.G. Summers, C.G. Smith, T.S. Maughan, V. Escott-Price, J.P. Cheadle

Writing, review, and/or revision of the manuscript: M.G. Summers, C.G. Smith, T.S. Maughan, R. Kaplan, V. Escott-Price, J.P. Cheadle

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M.G. Summers

Study supervision: T.S. Maughan, J.P. Cheadle

Published OnlineFirst November 4, 2016; DOI: 10.1158/1078-0432.CCR-16-1541

Other (responsible for data analyses and drafting of paper): M.G. Summers
Other (CI of associated study): T.S. Maughan
Other (co-directed this study): V. Escott-Price, J.P. Cheadle

Acknowledgments

We thank the patients and their families who participated and gave their consent for this research and the investigators and pathologists throughout the United Kingdom who submitted samples for assessment.

Grant Support

This work was supported by the Wales Gene Park, Cancer Research Wales, and the National Institute for Social Care and Health Research Cancer Genetics Biomedical Research Unit. The COIN and COIN-B trials were funded by Cancer Research UK and an educational grant from Merck-Serono. COIN and COIN-B were coordinated by the Medical Research Council Clinical Trials Unit and conducted with the support of the National Institute of Health Research Cancer Research Network.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 17, 2016; revised October 14, 2016; accepted October 31, 2016; published OnlineFirst November 4, 2016.

References

1. Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer 2009;9:489–99.
2. Köhne CH, Cunningham D, Di Costanzo F, Glimelius B, Blijham G, Aranda E, et al. Clinical determinants of survival in patients with 5-fluorouracil-based treatment for metastatic colorectal cancer: results of a multivariate analysis of 3825 patients. Ann Oncol 2002;13:308–17.
3. Haydon AM, Macinnis RJ, English DR, Giles GG. Effect of physical activity and body size on survival after diagnosis with colorectal cancer. Gut 2006;55:62–7.
4. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ 2007;335:1134.
5. Leitch EF, Chakrabarti M, Crozier JE, McKee RF, Anderson JH, Horgan PG, et al. Comparison of the prognostic value of selected markers of the systematic inflammatory response in patients with colorectal cancer. Br J Cancer 2007;97:1266–70.
6. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density and location of immune cells with human colorectal tumors predict clinical outcome. Science 2006;313:1960–4.
7. Smith CG, Fisher D, Harris R, Maughan TS, Phipps AI, Richman SD, et al. Analyses of 7,635 patients with colorectal cancer using independent training and validation cohorts show that rs9929218 in CDH1 is a prognostic marker of survival. Clin Can Res 2015;21:3453–61.
8. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol 2005;23:609–18.
9. Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut 2008;57:941–50.
10. Lochhead P, Kuchiba A, Imamura Y, Liao X, Yamauchi M, Nishihara R, et al. Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. J Natl Can Inst 2013;105:1151–6.
11. Eklöf V, Wikberg ML, Edin S, Dahlin AM, Johnsson BA, Oberg A, et al. The prognostic role of KRAS, BRAF, PIK3CA and PTEN in colorectal cancer. Br J Cancer 2013;108:2153–63.
12. Maughan TS, Adams RA, Smith CG, Meade AM, Seymour MT, Wilson RH, et al. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. The Lancet 2011;377:2103–14.
13. Smith CG, Fisher D, Claes B, Maughan TS, Idziaszczyk S, Peuteman G, et al. Somatic profiling of the epidermal growth factor receptor pathway in tumours from patients with advanced colorectal cancer treated with chemotherapy ± cetuximab. Clin Can Res 2013;19:4104–13.
14. Wasan H, Meade AM, Adams R, Wilson R, Pugh C, Fisher D, et al. Intermittent chemotherapy plus either intermittent or continuous cetuximab for first-line treatment of patients with KRAS wild-type advanced colorectal cancer (COIN-B): a randomised phase 2 trial. Lancet Oncol 2014;15:631–9.
15. Cremolini C, Di Bartolomeo M, Amatu A, Antoniotti C, Moretto R, Berenato R, et al. BRAF codons 594 and 596 mutations identify a new molecular subtype of metastatic colorectal cancer at favorable prognosis. Ann Oncol 2015;26:2092–7.
16. Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, et al. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell 2004;116:855–67.
17. Ikenoue T, Hikiba Y, Kanai F, Tanaka Y, Imamura J, Imamura T, et al. Functional analysis of mutations within the kinase activation segment of B-Raf in human colorectal tumors. Can Res 2003;63:8132–7.
18. Schirripa M, Cremolini C, Loupakis F, Morvillo M, Bergamo F, Zoratto F, et al. Role of NRAS mutations as prognostic and predictive markers in metastatic colorectal cancer. Int J Cancer 2015;136:83–90.
19. Burd CE, Liu W, Huynh MV, Waqas MA, Gillahan JE, Clark KS, et al. Mutation-specific RAS oncogenicity explains NRAS codon 61 selection in melanoma. Can Discov 2014;4:1418–29.
20. Yamauchi M, Morikawa T, Kuchiba A, Imamura Y, Qian ZR, Nishihara R, et al. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. Gut 2012;61:847–54.
21. Rosty C, Young JP, Walsh MD, Clendenning M, Walters RJ, Pearson S, et al. Colorectal carcinomas with KRAS mutation are associated with distinctive morphological and molecular features. Mod Pathol 2013;26:825–34.
22. Tran B, Kopetz S, Tie J, Gibbs P, Jiang ZQ, Lieu CH, et al. Impact of BRAF mutation and microsatellite instability on the pattern of metastatic spread and prognosis in metastatic colorectal cancer. Cancer 2011;117:4623–32.

OXFORD

JNCI J Natl Cancer Inst (2019) 111(8): djy215

doi: 10.1093/jnci/djy215
First published online January 14, 2019
Article

ARTICLE

Pattern Recognition Receptor Polymorphisms as Predictors of Oxaliplatin Benefit in Colorectal Cancer

Victoria Gray, Sarah Briggs, Claire Palles, Emma Jaeger, Timothy Iveson, Rachel Kerr, Mark P. Saunders, James Paul, Andrea Harkin, John McQueen, Matthew G. Summers, Elaine Johnstone, Haitao Wang, Laura Gatcombe, Timothy S. Maughan, Richard Kaplan, Valentina Escott-Price, Nada A. Al-Tassan, Brian F. Meyer, Salma M. Wakil, Richard S. Houlston, Jeremy P. Cheadle, Ian Tomlinson, David N. Church

See the Notes section for the full list of authors' affiliations.
Correspondence to: David N. Church, DPhil, Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK (e-mail: dchurch@well.ox.ac.uk).

Abstract

Background: Constitutional loss of function (LOF) single nucleotide polymorphisms (SNPs) in pattern recognition receptors *FPR1*, *TLR3*, and *TLR4* have previously been reported to predict oxaliplatin benefit in colorectal cancer. Confirmation of this association could substantially improve patient stratification.
Methods: We performed a retrospective biomarker analysis of the Short Course in Oncology Therapy (SCOT) and COIN/COIN-B trials. Participant status for LOF variants in *FPR1* (rs867228), *TLR3* (rs3775291), and *TLR4* (rs4986790/rs4986791) was determined by genotyping array or genotype imputation. Associations between LOF variants and disease-free survival (DFS) and overall survival (OS) were analyzed by Cox regression, adjusted for confounders, using additive, dominant, and recessive genetic models. All statistical tests were two-sided.
Results: Our validation study populations included 2929 and 1948 patients in the SCOT and COIN/COIN-B cohorts, respectively, of whom 2728 and 1672 patients had functional status of all three SNPs determined. We found no evidence of an association between any SNP and DFS in the SCOT cohort, or with OS in either cohort, irrespective of the type of model used. This included models for which an association was previously reported for rs867228 (recessive model, multivariable-adjusted hazard ratio [HR] for DFS in SCOT = 1.19, 95% confidence interval [CI] = 0.99 to 1.45, $P = .07$; HR for OS in COIN/COIN-B = 0.92, 95% CI = 0.63 to 1.34, $P = .66$), and rs4986790 (dominant model, multivariable-adjusted HR for DFS in SCOT = 0.86, 95% CI = 0.65 to 1.13, $P = .27$; HR for OS in COIN/COIN-B = 1.08, 95% CI = 0.90 to 1.31, $P = .40$).
Conclusion: In this prespecified analysis of two large clinical trials, we found no evidence that constitutional LOF SNPs in *FPR1*, *TLR3*, or *TLR4* are associated with differential benefit from oxaliplatin. Our results suggest these SNPs are unlikely to be clinically useful biomarkers.

The antitumor immune response is an important determinant of clinical outcome in colorectal cancer (CRC). To date, attention has primarily focused on the role of the adaptive immune system, and particularly the T-cell response, the increasing intensity of which correlates with reduced recurrence in early-stage CRC (1,2). Although the influence of the innate immune system to clinical outcome is less well understood, several studies have suggested that this may also exert a meaningful antitumor effect through the recognition of endogenous ligands presented by dying cells (3–7). This effect has been reported to be especially relevant in the context of cell death induced by anthracyclines and oxaliplatin (3–5), an analog of cisplatin used

Received: June 10, 2018; **Revised:** August 22, 2018; **Accepted:** November 19, 2018

commonly in the systemic therapy of CRC (8). Pattern recognition receptors present endogenous ligands to macrophages and as such are essential components of the innate immune response (9). Constitutional variants in several genes encoding these proteins have been shown to alter the innate immune response to systemic infection (10). Recently, polymorphisms that result in putative loss of function (LOF) alterations in pattern recognition receptor genes have also been reported to influence benefit from anthracycline and oxaliplatin chemotherapy (4–7). These variants, which affect *FPR1* [rs867228: c.1037A>C, p.Glu346Ala, where Ala is the LOF allele (11)], *TLR3* [rs3775291: c.1234C>T, p.Leu412Phe, where Phe is the LOF allele (12)], and *TLR4* [rs4986790: c.896A>G, p. Asp299Gly, where Gly is the LOF allele (4,7)] in strong linkage disequilibrium with rs4986791: c.1196C>T, p.Thr399Ile, are proposed to act by attenuating the immune response against the immunogenic cell death caused by these agents (4–7). These associations were reflected in statistically significant differences in both progression-free and overall survival (OS) between patients bearing LOF and functional alleles in these genes when treated with these agents (hazard ratios [HRs] for LOF allele of 1.37–2.13; summarized in Supplementary Table 1, available online) (4–7,13,14). If validated, these variants could be used as biomarkers to target these toxic therapies to those most likely to benefit from them, resulting in less harm to patients and cost savings for healthcare providers. Because anthracyclines and oxaliplatin are the mainstays of systemic treatment against two common cancers (breast and colorectal, respectively) (15,16) and because these LOF polymorphisms are relatively common (prevalence of 5% to 80% in populations of European descent), confirmation of this association could affect many thousands of patients each year in Europe and the United States alone. The purpose of this validation study was to confirm this association in the context of oxaliplatin treatment for CRC by analysis of two well-defined, prospectively treated cohorts from the Short Course in Oncology Therapy (SCOT) and COIN/COIN-B trials (17,18), encompassing both early-stage and advanced disease.

Methods

Clinical Trials

Details of the SCOT (ISRCTN59757862), COIN (ISRCTN27286448), and COIN-B (ISRCTN38375681) trials have been published previously (17–20). Briefly, the SCOT trial compared the efficacy of 12 weeks of oxaliplatin-based adjuvant chemotherapy with the previous standard of care of 24 weeks of treatment in high-risk, stage II (defined as one or more of: pT4 primary tumor, tumor obstruction, fewer than 10 lymph nodes harvested, grade 3 histology, perineural invasion, or extramural venous or lymphatic vascular invasion), or stage III colon or rectal cancer. The trial randomized 6088 patients between March 2008 and November 2013, of whom 6065 consented for their data to be used for the intention to treat analyses. At its primary analysis, the attenuated course of chemotherapy was confirmed to be noninferior to the standard of care (HR = 1.01, 95% CI = 0.91 to 1.11, test for noninferiority P = .012) (17). As part of the study, participants at selected centers were invited to participate in a translational substudy, the TransSCOT study. Tissue and blood samples were collected from these patients and constitutional DNA was extracted for translational studies. Following informed consent, 3109 patients provided samples for analysis. The COIN trial examined both the efficacy of the anti-EGFR monoclonal antibody cetuximab added to oxaliplatin-based chemotherapy and the impact of interrupting treatment in patients with stable or responding metastatic CRC after 12 to 16 weeks of systemic therapy (18). The trial recruited 2445 patients between March 2005 and May 2008. At its primary analysis, no statistically significant difference was observed between the chemotherapy-only and the chemotherapy plus cetuximab groups (20), and the comparison between intermittent and continuous chemotherapy failed to confirm noninferiority of interrupting treatment (18). The COIN-B study compared intermittent chemotherapy with either intermittent or continuous cetuximab in 226 patients with metastatic CRC (19). Among 169 patients with *KRAS* wild-type disease, analysis suggested greater activity of continuous cetuximab, though this difference was not statistically significant. As part of ancillary translational studies, 2244 study participants in COIN and COIN-B donated blood samples for DNA extraction and analysis. Given their similar patient populations and treatments (21), the COIN and COIN-B biomarker cohorts were combined for all analyses.

DNA Extraction, Genotyping, and Imputation

DNA was extracted from EDTA-venous bloods using standard methods. After exclusion of samples that failed DNA extraction (n = 28) and those for which trial IDs were missing or duplicates (n = 14), 3067 DNA samples from the SCOT cohort were genotyped using the Global Screening Array (Illumina, San Diego, CA). Genotyping quality control entailed removal of any sample or single nucleotide polymorphism (SNP) with more than 2% missing data, any sample with an outlying heterozygosity rate, any sample with discordant reported sex and genotype imputed sex, and any SNP violating Hardy-Weinberg equilibrium at *P* less than 1×10^{-10} (n = 66 samples removed; n = 32 850 SNPs removed). Identity by descent analysis was conducted in PLINK 1.9 (22) and population stratification was examined using EIGENSTRAT (23). Related individuals (n = 8) were removed (IBD > 0.185) along with those with non-European ancestry (n = 54, as assessed by merging SCOT with HapMap release 23a and removing outliers based on eigenvector 1). Genotypes for 2939 remaining individuals were phased using SHAPEIT (24) and imputed using IMPUTE2 (25) and the UK10K + 1000 genomes merged reference panel. Of the SNPs analyzed in this study, rs3775291, rs4986790, and rs4986791 were directly genotyped. The fourth, rs867228, was imputed with an info score of 0.95. For this imputed SNP, genotype probabilities were converted to genotypes using gtool (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html) with a minimum probability threshold of .9 set for specifying per sample genotypes.

Cases from the COIN and COIN-B studies were genotyped using Affymetrix Axiom Arrays according to the manufacturer's recommendations (Affymetrix, Santa Clara, CA) at the King Faisal Specialist Hospital and Research Center, Saudi Arabia (under IRB approval 2110033). We excluded individuals from analysis if they failed one or more of the following thresholds: overall successfully genotyped SNPs less than 95% (n = 122), discordant sex information (n = 8), classed as out of bounds by Affymetrix (n = 30), duplication or cryptic relatedness (identity by descent >0.185, n = 4), and evidence of non-white European ancestry by principal components analysis-based analysis in comparison with HapMap samples (n = 130). Imputation was performed using 1000 Genomes Project Pilot data as a reference

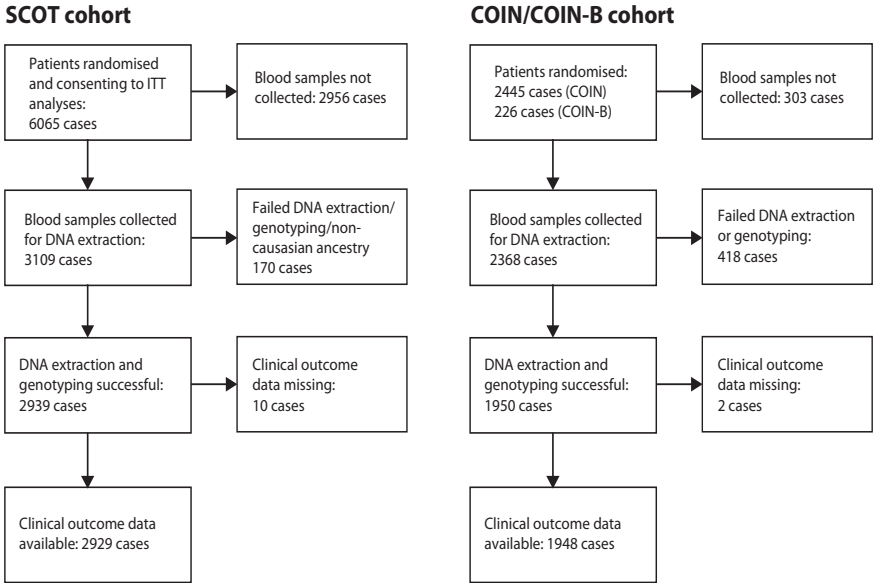

Figure 1. CONSORT diagram showing flow of patients analyzed in the study. ITT = intention to treat.

panel (26). Genetic linkage of SNPs was determined by calculation of D' and R2 using PLINK 1.9 (22).

Statistical Analyses

Comparison between groups was made using unpaired Student *t* test for continuous variables (eg, age) and either χ^2 or Fisher exact test for categorical variables (eg, mutation present vs absent, responder vs nonresponder). Biomarker analyses in this study were performed and are reported in accordance with the REMARK guidelines (27). All analyses were prespecified and are detailed in Supplementary Table 2 (available online). Survival endpoints included disease-free survival (DFS, defined as time from study randomization to CRC recurrence or death from any cause in SCOT only) and OS (defined as time from randomization to death from any cause in both cohorts). Progression-free survival was not used as an endpoint in the COIN/COIN-B trials in view of the difficulty in defining its duration in the context of intermittent chemotherapy, which was tested in both studies. Survival curves for SNP genotypes were plotted using the Kaplan-Meier method and analyzed by the log-rank test. Survival endpoints were also analyzed by univariate and multivariable Cox proportional hazards models, under additive, recessive, and dominant genetic models (eg, for rs867228, which has alleles A and C—of which C is the LOF allele—the additive model implies CC [2] vs CA [1] vs AA [0], modelled as a continuous variable; the recessive model implies CC [1] vs both CA and AA [0]; and the dominant model implies both CC and CA [1] vs AA). Proportionality of hazards was confirmed by inspection of scaled Schoenfeld residuals. For the multivariable analyses, adjustment was made for baseline demographic variables (age, sex), clinicopathological and molecular covariables of known prognostic value where available, and treatment type and schedule depending on the cohort. In the SCOT analyses, these comprised age, sex, disease site (colon vs rectum), primary tumor stage (pT1–2 vs pT3 vs pT4), nodal status (N0 vs N1 vs N2), treatment regimen (FOLFOX or CAPOX), and treatment duration (24 vs 12 weeks). In the COIN/COIN-B analyses, these comprised age, sex, disease site (colon vs rectum), World Health Organization (WHO) performance status (0 or 1 vs 2), primary tumor resection (unresected vs resected), tumor *KRAS*, *NRAS*, and *BRAF* mutation status (mutated vs wild type), patient white blood cell count (<10 000 cells per μL vs ≥10 000 cells per μL), cetuximab treatment (yes vs no), chemotherapy regimen (FOLFOX vs CAPOX), and chemotherapy schedule (intermittent vs continuous). In both cases, covariables were prespecified and no selection procedure (eg, backwards elimination) was performed. Models included all cases for which data were available and excluded those with missing data. *P* values for individual predictors in Cox models were calculated by the Wald test. Statistical analyses were performed in R version 3.4.4 (CRAN Corporation) and STATA version 13 (StataCorp, College Station, TX). All statistical tests were two-sided. Statistical significance was accepted at *P* less than .05. No correction for multiple testing was applied.

Ethical Approval

Informed consent for the collection and analysis of samples was provided by study participants at the time of study recruitment under trial-specific ethical approval. Molecular analysis of samples from the SCOT cohort was performed under North West – Liverpool Central Research Committee approval (17/NW/0252). Molecular analysis of COIN/COIN-B samples was performed under REC approval (04/MRE06/60).

Results

Patient Characteristics and SNP Genotyping

The CONSORT diagram demonstrating the flow of patients eligible for this biomarker study is shown in Figure 1. Demographic

Table 1. Baseline characteristics of SCOT and combined COIN/COIN-B cohorts

Variable	SCOT No. (%)	COIN and COIN-B No. (%)	P
Total	2929 (100)	1948 (100)	—
Median age, y (range)	65 (23–84)	53 (18–87)	<.001*
Sex			
Male	1795 (61.3)	1270 (65.2)	<.001†
Female	1134 (38.7)	678 (34.8)	
Unknown	0 (0.0)	0 (0.0)	
Disease stage			
II	585 (20.0)	0 (0.0)	—
III	2344 (80.0)	0 (0.0)	
IV	0 (0.0)	1948 (100.0)	
Unknown	0 (0.0)	0 (0.0)	
Primary tumor stage			
pT1	94 (3.2)	NA	—
pT2	285 (9.7)	NA	
pT3	1694 (57.8)	NA	
pT4	856 (29.2)	NA	
Unknown	0 (0.0)	NA	
Nodal stage			
N0	585 (20.0)	NA	—
N1	1695 (57.9)	NA	
N2	649 (22.2)	NA	
Unknown	0 (0.0)	NA	
Primary tumor location			
Colon	2346 (80.1)	1325 (67.9)	<.001†
Rectum	583 (19.9)	621 (31.9)	
Unknown	0 (0.0)	2 (0.2)	
Primary tumor resected			
No	0 (0.0)	821 (42.1)	—
Yes	2929 (100.0)	1127 (57.9)	
Unknown	0 (0.0)	0 (0.0)	
Peritoneal metastases			
No	NA	1519 (78.0)	—
Yes	NA	259 (13.3)	
Unknown	NA	170 (8.7)	
KRAS mutation status			
Wild-type	ND	989 (50.8)	—
Mutant	ND	636 (32.6)	
Unknown	ND	323 (16.6)	
NRAS mutation status			
Wild type	ND	1506 (77.3)	—
Mutant	ND	69 (3.6)	
Unknown	ND	373 (19.1)	
BRAF mutation status			
Wild type	ND	1438 (73.8)	—
Mutant	ND	143 (7.3)	
Unknown	ND	367 (18.9)	
FPR1 rs867228 genotype			
AA	116 (4.0)	49 (2.5)	.003†
AC	813 (27.8)	444 (22.8)	
CC	1799 (61.4)	1179 (60.5)	
Unknown	201 (6.9)	276 (14.2)	
TLR3 rs3775291 genotype			
CC	1486 (50.7)	934 (47.9)	.005†
CT	1207 (41.2)	810 (41.6)	
TT	231 (7.9)	204 (10.5)	
Unknown	5 (0.2)	0 (0.0)	
TLR4 rs4986790 genotype			
AA	2581 (88.1)	1744 (89.5)	.11†
AG	333 (11.4)	200 (10.3)	
GG	15 (0.5)	4 (0.2)	
Unknown	0 (0.0)	0 (0.0)	
TLR4 rs4986791 genotype			
CC	2568 (90.7)	1726 (88.6)	.12†
CT	344 (11.7)	218 (11.2)	
TT	17 (0.6)	4 (0.2)	
Unknown	0 (0.0)	0 (0.0)	

*Determined by two-sided unpaired Student *t* test. NA = not applicable; ND = not determined; pT = pathological tumor (T) stage; SCOT = Short Course in Oncology Therapy.
†Determined by two-sided χ^2 test or Fisher exact test in the case of rs4986791 (in cases of SNP genotypes, values are calculated from cases in which SNP status was determined).

and clinicopathological characteristics of the 2929 SCOT cases with samples informative for this analysis were broadly similar to those of the SCOT trial population as a whole, although they differed statistically significantly, albeit modestly, from the nonbiomarker population in age, disease site, disease stage, and nodal status (Supplementary Table 2, available online). Characteristics of 2244 patients in the COIN/COIN-B biomarker subgroup were similar to the combined COIN/COIN-B trial population (not shown). Details of baseline demographic, clinicopathological, and molecular variables, and SNP genotypes in cases from both biomarker cohorts are provided in Table 1. Of 2929 patients in the SCOT cohort, 2728 (93.1%), 2924 (99.9%), and 2929 (100%) underwent successful genotyping or imputation and were informative for analysis of rs867228, rs3775291, and rs4986790/rs4986791 respectively. The slightly lower number of cases informative for rs867228 reflects the exclusion of those in which the genotype could not be imputed with high confidence. The corresponding numbers in the COIN/COIN-B cohort of 1948 patients were 1672 (85.6%), 1948 (100%), and 1948 (100%) respectively. The allelic frequencies of all SNPs in both cohorts were concordant with the reported population frequency in ExAC (28), EVS (29), and UK10K (30). As expected, rs4986790 and rs4986791 were in strong linkage disequilibrium in both the SCOT ($D' = 0.99$ and $r^2 = 0.93$) and COIN/COIN-B ($D' = 0.99$ and $r^2 = 0.89$) cohorts. Because analyses of these two SNPs individually yielded essentially identical results (Supplementary Figure 1, available online), we largely limited subsequent investigations to rs4986790.

The effect sizes (hazard ratios) of each SNP detectable in multivariable analyses using recessive and dominant genetic models, based on a power ($1-\beta$) of 0.8 and a two-sided α of 0.05, are shown for both cohorts in Supplementary Table 4 (available online). For comparison with previous reports, our power to detect an association of identical effect size using the same (recessive) model to that previously reported for the *FPR1* rs867228 SNP was 1.0 and 0.995 for DFS and OS, respectively, in the SCOT cohort and 1.0 for OS in the COIN/COIN-B cohort. Our power to detect an association of the same effect size as that previously reported for the *TLR4* rs4986790 SNP using the same (dominant) model was 0.65 and 0.31 for DFS and OS, respectively, in the SCOT cohort and 0.96 for OS in the COIN/COIN-B cohort.

Pattern Recognition SNPs and Clinical Outcome in the SCOT Cohort

Biomarker analyses were performed with data used for the primary analysis of the SCOT trial, at which point the 2929 patients

Table 2. Univariate and multivariable analyses of DFS and OS in SCOT cohort by LOF SNP*

			Univariate analysis				Multivariable analysis			
			DFS		OS		DFS		OS	
Polymorphism/genetic model	No.	DFS events	OS events	HR (95% CI)	P†	HR (95% CI)	P†	HR (95% CI)	P†	HR (95% CI)	P†
rs867228 (FPR1 c.1037A>C)	2728	487	167								
Additive	—	—	—	1.13 (0.96 to 1.32)	.15	1.09 (0.83 to 1.44)	.53	1.16 (0.98 to 1.37)	.08	1.10 (0.84 to 1.47)	.48
Recessive	—	—	—	1.15 (0.95 to 1.40)	.15	1.07 (0.77 to 1.49)	.67	1.19 (0.99 to 1.45)	.07	1.10 (0.79 to 1.53)	.56
Dominant	—	—	—	1.17 (0.73 to 1.88)	.50	1.40 (0.57 to 3.41)	.46	1.16 (0.73 to 1.87)	.53	1.32 (0.54 to 3.22)	.54
rs3775291 (TLR3 c.1234C>T)	2924	536	186								
Additive	—	—	—	1.05 (0.92 to 1.19)	.52	1.15 (0.93 to 1.44)	.29	1.02 (0.90 to 1.17)	.68	1.13 (0.91 to 1.41)	.27
Recessive	—	—	—	1.24 (0.92 to 1.66)	.15	1.46 (0.92 to 2.32)	.11	1.14 (0.85 to 1.52)	.38	1.32 (0.83 to 2.10)	.24
Dominant	—	—	—	1.01 (0.85 to 1.19)	.95	1.12 (0.84 to 1.49)	.44	1.00 (0.85 to 1.19)	.97	1.12 (0.84 to 1.49)	.44
rs4986790 (TLR4 c.896A>G)	2929	538	186								
Additive	—	—	—	0.92 (0.71 to 1.19)	.52	0.89 (0.57 to 1.39)	.62	0.89 (0.69 to 1.16)	.39	0.87 (0.56 to 1.36)	.54
Recessive	—	—	—	1.49 (0.55 to 4.00)	.42	1.93 (0.48 to 7.76)	.36	1.58 (0.59 to 4.25)	.36	1.82 (0.44 to 7.40)	.40
Dominant	—	—	—	0.89 (0.67 to 1.17)	.40	0.83 (0.51 to 1.35)	.45	0.86 (0.65 to 1.13)	.27	0.81 (0.50 to 1.32)	.39

*Both univariate and multivariable analyses use all informative cases. Hazard ratios show risk associated with reported LOF allele (underscored) for each SNP as follows: rs867228: *FPR1* c.1037A>C p.Glu346Ala; rs3775291: *TLR3* c.1234C>T, p.Leu412Phe; rs4986790: *TLR4* c.896A>G, p. Asp299Gly. Corresponding associations from rs4986791 (*TLR4* c.1196C>T, p.Thr399Ile), which is tightly linked to rs4986790, were essentially identical to those obtained from analysis of rs4986790 and are not shown. Multivariable-adjusted HRs were adjusted for age, sex, disease site (colon vs rectum), primary tumor stage (pT1–2 vs pT3 vs pT4), nodal status (N0 vs N1 vs N2), treatment regimen (FOLFOX or CAPOX), and treatment duration (24 vs 12 weeks). Prognostic associations of covariables are shown in Supplementary Table 5 (available online). CI = confidence interval; DFS = disease-free survival; HR = hazard ratio; LOF = loss of function; OS = overall survival; pT = pathological tumor (T) stage; SCOT = Short Course in Oncology Therapy.
†P values were calculated by two-sided Wald test.

in the biomarker cohort had a median follow-up of 36.8 months, and 538 DFS events and 186 deaths had occurred (Table 2). Comparing survival curves by the log-rank test, univariate and multivariable Cox models demonstrated no statistically significant association of any SNP irrespective of genetic model imposed (Figure 2, Table 2, details of covariables in multivariable models provided in Supplementary Table 5, available online). This included models for which an association was previously reported for rs867228 (5) (recessive model, multivariable-adjusted HR for DFS = 1.19, 95% CI = 0.99 to 1.45, P = .07) and rs4986790 (4) (dominant model, multivariable-adjusted HR for DFS = 0.86, 95% CI = 0.65 to 1.13, P = .27) (Table 2, Supplementary Table 5, available online).

A previous study reported that the association of the *FPR1* LOF polymorphism rs867228 was only evident in patients with functional *TLR3* or *TLR4*, consistent with their participation in the same pathway (5). We therefore examined this in the SCOT biomarker cohort after stratifying by *TLR3* (rs3775291) and *TLR4* (rs4986790) status. These analyses did not confirm the previously reported, statistically significant association with DFS in the context of either functional *TLR3* background (multivariable-adjusted HR for additive model = 1.02, 95% CI = 0.82 to 1.27, P = .85; recessive model HR = 1.01, 95% CI = 0.78 to 1.31, P = .91; dominant model HR = 1.09, 95% CI = 0.59 to 2.00, P = .78) or functional *TLR4* background (additive model HR = 1.17, 95% CI = 0.99 to 1.40, P = .07; recessive model HR = 1.20, 95% CI = 0.98 to 1.48, P = .08; dominant model HR = 1.31, 95% CI = 0.79 to 1.20, P = .30). Similarly, no statistically significant association of rs867228 with DFS was observed in cases with functional polymorphisms at both of these loci (multivariable-adjusted HR for additive model = 0.97, 95% CI = 0.77 to 1.21, P = .76; recessive model HR = 0.92, 95% CI = 0.70 to 1.22, P = .58; dominant model HR = 1.14, 95% CI = 0.60 to 2.16, P = .68) (Supplementary Figure 2, available online).

Pattern Recognition SNPs, Clinical Outcome, and Oxaliplatin Response in the COIN/COIN-B Cohort

Corresponding analyses were performed on the COIN/COIN-B cohort in which the median follow-up of the 1948 patients was 23.2 months, by which time 1453 deaths had occurred. Similar to the SCOT analyses, there was no statistically significant association of either SNP with OS by either log-rank test or univariate or multivariable Cox regression, regardless of model (Figure 3, Table 3, details of covariables in multivariable models provided in Supplementary Table 6, available online). Again, this included the recessive model for rs867228 (5) (multivariable-adjusted HR for OS = 0.92, 95%CI = 0.63 to 1.34, P = .66), and the dominant model for rs4986790 (4) (multivariable-adjusted HR for OS = 1.08, 95% CI = 0.90 to 1.31, P = .40) (Table 3, Supplementary Table 6, available online). Likewise, prespecified subgroup analyses stratified by *TLR3* and *TLR4* status revealed no evidence of an association between *FPR1* status and OS in the context of functional *TLR3* (multivariable-adjusted HR for additive model = 0.93, 95% CI = 0.78 to 1.10, P = .37; recessive model HR = 0.91, 95% CI = 0.56 to 1.48, P = .71; dominant model HR = 0.91, 95% CI = 0.74 to 1.12, P = .36), or functional *TLR4* (additive model HR = 1.03, 95% CI = 0.90 to 1.17, P = .66; recessive model HR = 1.00, 95% CI = 0.68 to 1.49, P = .99; dominant model HR = 1.04, 95% CI = 0.89 to 1.21, P = .62). Similar to the results from the SCOT cohort, no statistically significant association was observed in cases with functional polymorphisms at both loci (multivariable-adjusted HR for additive model = 0.99, 95% CI = 0.82 to 1.19, P = .87; recessive model = 1.22, 95% CI = 0.73 to 2.04, P = .43; dominant model HR = 0.95, 95% CI = 0.76 to 1.18, P = .63) (Supplementary Figure 3, available online).

An additional analysis according to radiological response to oxaliplatin-based chemotherapy after 12 weeks of therapy

ARTICLE

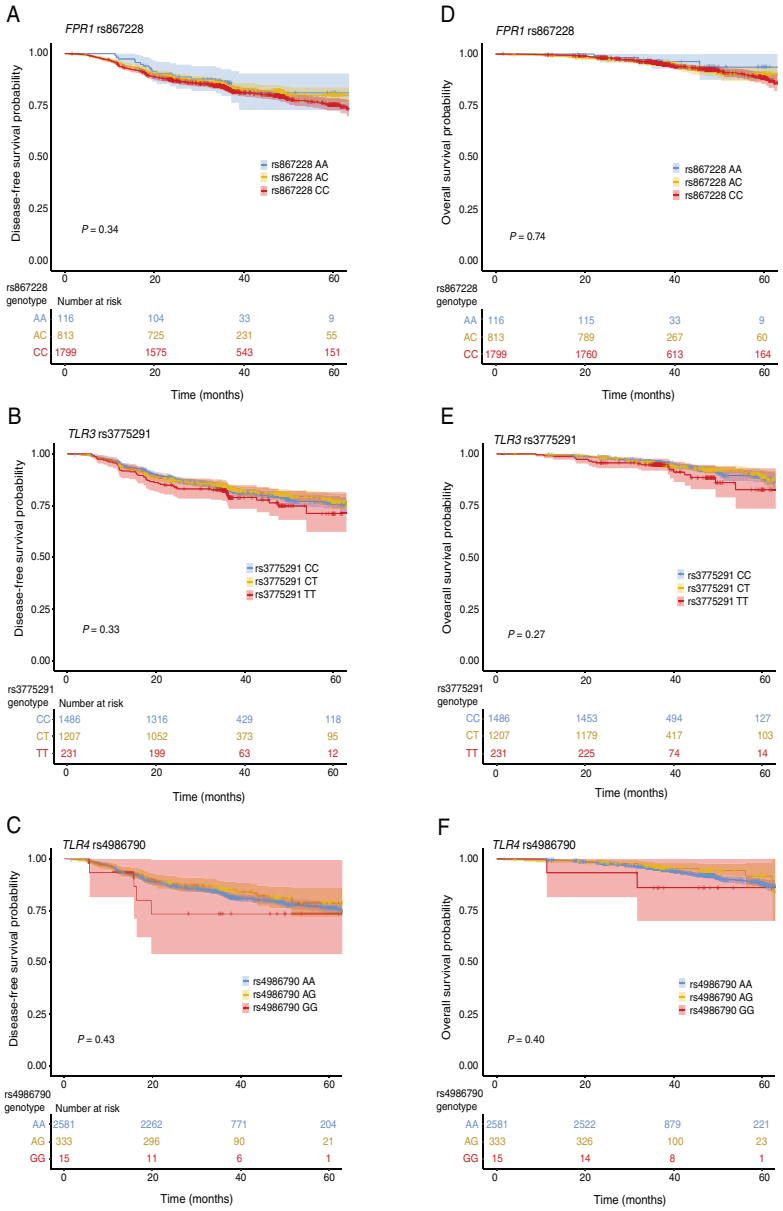

Figure 2. *FPR1*, *TLR3*, and *TLR4* loss of function (LOF) single nucleotide polymorphisms (SNPs), disease-free survival (DFS), and overall survival (OS) in Short Course in Oncology Therapy (SCOT) cohort. Kaplan Meier curves showing DFS for patients in SCOT cohort by pattern recognition receptor SNPs rs867228 (*FPR1* c.1037A>C p.Glu346Ala) (**A**), rs3775291 (*TLR3* c.1234C>T p.Leu412Phe) (**B**), and rs4986790 (*TLR4* c.896A>G, p.Asp299Gly) (**C**) (LOF allele/amino acid underscored in each case). Corresponding results for OS are shown in **D–F**. Analyses of the rs4987691 (*TLR4* c.1196C>T, p. Thr399Ile) polymorphism, which is strongly linked with rs4986790, were essentially identical to **C** and **F** and are provided as Supplementary Figure 1 (available online). Shaded areas represent 95% confidence intervals. *P* values indicate comparison of all groups by the two-sided log-rank test.

[complete or partial response vs stable or progressive disease by RECIST 1.0 (31)] revealed no difference in the proportions of functional and LOF alleles between responders and nonresponders for rs867228 ($P = .90$, χ^2 test), rs3775291 ($P = .68$, χ^2 test), or rs4986790 ($P = .64$, Fisher exact test).

Discussion

Previous studies have suggested that LOF polymorphisms in the pattern recognition receptors *FPR1* (rs867228), *TLR3* (rs3775291), and *TLR4* (rs4986790/rs4986791) decrease the presentation of

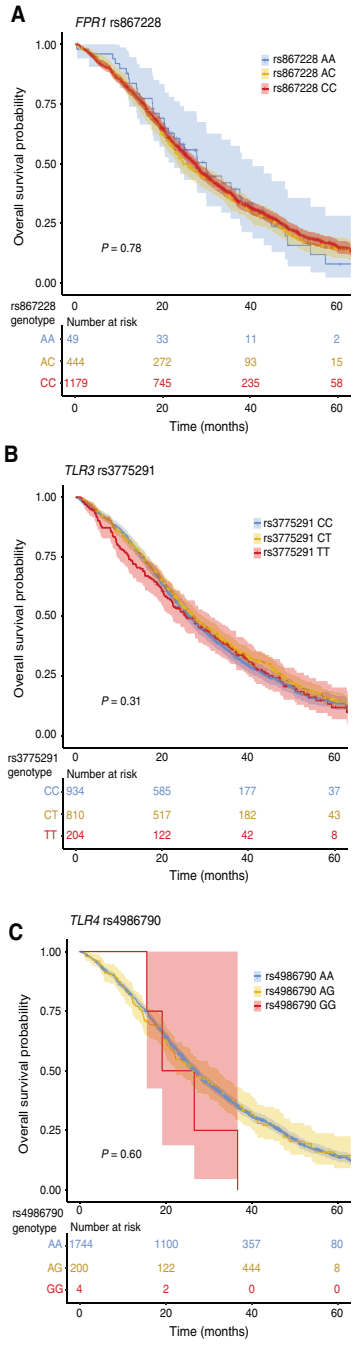

Figure 3. *FPR1*, *TLR3*, and *TLR4* loss of function (LOF) single nucleotide polymorphisms (SNPs) and overall survival (OS) in COIN/COIN-B cohort. Kaplan Meier curves showing OS for patients in combined COIN/COIN-B cohort by pattern recognition receptor SNPs rs867228 (*FPR1* c.1037A>C p.Glu346Ala) (**A**), rs3775291 (*TLR3* c.1234C>T p.Leu412Phe) (**B**), and rs4986790 (*TLR4* c.896A>G, p.Asp299Gly) (**C**) (LOF allele/amino acid underscored in each case). Analyses of the rs4987691 (*TLR4* c.1196C>T, p.Thr399Ile) polymorphism, which is strongly linked with rs4986790, were essentially identical to **C** and are not shown. Shaded areas represent 95% confidence intervals. *P* values indicate comparison of all groups by the two-sided log-rank test.

ligand to the innate immune system by dying cells (3–7). This, in turn, is proposed to reduce the efficacy of anthracycline and oxaliplatin chemotherapy, the activities of which depend in part on the induction of immunogenic cell death (3–5,7). In this study of nearly 5000 patients with CRC treated with oxaliplatin, we failed to confirm any of these associations. The 95% confidence intervals for the association of each SNP with DFS and OS in the SCOT cohort and OS in the COIN/COIN-B cohort all included the estimate of no effect. Although our data by no means exclude an immunomodulatory effect of these SNPs, they suggest that they are very unlikely to be clinically useful as predictive biomarkers for oxaliplatin benefit in CRC. The discordance between our results and those from previous studies may be explained by the increased risk of false-positive associations in the smaller cohorts they used, and in the case of rs867228, an apparent misclassification of the functional and LOF alleles in the survival analyses (the functional *FPR1* allele c.1037A, p.346Glu appeared to be incorrectly classified as LOF in all analyses in the study by Vacchelli et al.) (5). Our results underscore the importance of validation of encouraging findings from modestly sized studies in large, meticulously curated trial cohorts, even where preclinical data provide a plausible mechanism for an association.

Strengths of our study include its large size, defined clinical trial cohorts, standardized therapy, comprehensively annotated clinicopathological variables, and, in the case of the COIN/COIN-B cohort, molecular variables and mature outcome data. Consequently, our analyses were powered to detect even a modest association of most SNPs with clinical outcome and had a power of greater than 0.95 to detect an association of similar strength to that previously reported for the rs867228 and rs4986790 LOF variants (4,6). Limitations include the lack of molecular profiling in the SCOT trial, which meant that we were unable to test for an association of the SNPs with clinical outcome in specific tumor subgroups such as those with enhanced immunogenicity due to defective DNA mismatch repair or *POLE* exonuclease domain mutation.

In summary, in this study of two large clinical trial cohorts, we find no evidence that LOF SNPs in the pattern recognition receptors *FPR1*, *TLR3*, and *TLR4* are associated with differential benefit from oxaliplatin in CRC. Future studies may better define the complex relationship between cytotoxic therapeutic-induced cell death, pattern recognition SNPs, and the innate immune system.

Funding

This work was supported by The Oxford NIHR Comprehensive Biomedical Research Centre, Cancer Research UK (C6199/A10417 and C399/A2291), the European Union Seventh Framework Programme (FP7/207– 2013) grant 258236 collaborative project SYSCOL, European Research Council project EVOCAN, and core funding to the Wellcome Trust Centre for Human Genetics from the Wellcome Trust (090532/Z/09/Z). The SCOT trial was supported by the Medical Research Council (transferred to NETSCC—Efficacy and Mechanism Evaluation; grant reference G0601705), the Swedish Cancer Society, and Cancer Research UK Core Clinical Trials Unit Funding (funding reference C6716/A9894). The TRIAL sponsor was NHS Greater Glasgow & Clyde and University of Glasgow (Eudract reference 2007–003957–10; ISRCTN number 23516549). COIN and COIN-B

Table 3. Univariate and multivariable analyses of OS in combined COIN/COIN-B cohort by LOF SNP*

Polymorphism/genetic model	Univariate analysis				Multivariable analysis			
	No.	OS events	HR (95% CI)	P†	No.	OS events	HR (95% CI)	P†
rs867228 (FPR1 c.1037A>C)	1672	1241	—	—	1336	970	—	—
Additive	—	—	1.03 (0.93 to 1.14)	.60	—	—	0.99 (0.88 to 1.12)	.91
Recessive	—	—	0.98 (0.71 to 1.37)	.93	—	—	0.92 (0.63 to 1.34)	.66
Dominant	—	—	1.04 (0.92 to 1.18)	.52	—	—	1.05 (0.90 to 1.23)	.53
rs3775291 (TLR3 c.1234C>T)	1948	1453	—	—	1563	1150	—	—
Additive	—	—	0.98 (0.91 to 1.06)	.67	—	—	0.97 (0.89 to 1.06)	.56
Recessive	—	—	1.07 (0.90 to 1.26)	.45	—	—	1.08 (0.90 to 1.31)	.41
Dominant	—	—	0.94 (0.86 to 1.05)	.31	—	—	0.93 (0.83 to 1.04)	.20
rs4986790 (TLR4 c.896A>G)	1948	1453	—	—	1563	1150	—	—
Additive	—	—	1.03 (0.88 to 1.21)	.71	—	—	1.10 (0.91 to 1.33)	.31
Recessive	—	—	1.65 (0.61 to 4.40)	.31	—	—	2.91 (0.93 to 9.12)	.07
Dominant	—	—	1.02 (0.86 to 1.20)	.81	—	—	1.08 (0.90 to 1.31)	.40

*Both univariate and multivariable analyses use all informative cases (ie, cases lacking covariable data were excluded from multivariable models). Hazard ratios show risk associated with reported LOF allele (underscored) for each SNP as follows: rs867228: *FPR1* c.1037A>C p.Glu346Ala; rs3775291: *TLR3* c.1234C>T, p.Leu412Phe; rs4986790: *TLR4* c.896A>G, p. Asp299Gly. Corresponding associations from rs4986791 (*TLR4* c.1196C>T, p.Thr399Ile), which is tightly linked to rs4986790, were essentially identical to those obtained from analysis of rs4986790 and are not shown. Multivariable-adjusted HRs are adjusted for age, sex, disease site (colon vs rectum), World Health Organization (WHO) performance status (0 or 1 vs 2), primary tumor resection (unresected vs resected), tumor *KRAS*, *NRAS*, and *BRAF* mutation status (mutated vs wild type), patient white blood cell count (<10 000 cells/μL vs ≥10 000 cells/μL), cetuximab treatment (yes vs no), chemotherapy regimen (FOLFOX vs CAPOX), and chemotherapy schedule (intermittent vs continuous). Prognostic associations of covariables are shown in Supplementary Table 6 (available online). CI = confidence interval; HR = hazard ratio; LOS = loss of function; OS = overall survival; pT = pathological tumor (T) stage; SNP = single nucleotide polymorphism.
†*P* values were calculated by two-sided Wald test.

were coordinated by the Medical Research Council Clinical Trials Unit and conducted with the support of the National Institutes of Health Research Cancer Research Network. COIN and COIN-B translational studies were supported by the Bobby Moore Fund from Cancer Research UK, Tenovus, the Kidani Trust, Cancer Research Wales, and the National Institute for Social Care and Health Research Cancer Genetics Biomedical Research Unit.

SB is funded by an MRC Clinical Research Training Fellowship. CP is funded by a University of Birmingham Fellowship. NAA, BFM, and SMW were funded and supported by KFSHRC. RSH is supported by Cancer Research UK. DNC is funded by a Health Foundation/Academy of Medical Sciences Clinician Scientist Fellowship.

The cost of open access publication was provided by core funding to the Wellcome Centre for Human Genetics from the Wellcome Trust (203141/Z/16/Z).

Notes

Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff, UK (VG, MGS, JPC); Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK (SB, EJ, LG, DNC); Cancer Genetics and Evolution Laboratory, Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham, UK (SB, EJ, IT); Gastrointestinal Cancer Genetics Laboratory, Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham, UK (CP); Southampton University Hospital NHS Foundation Trust, Southampton, UK (TI, TSM); Department of Oncology, Old Road Campus Research Building, University of Oxford, Oxford, UK (RK, EJ, HW); Christie Hospital NHS Foundation Trust, Manchester, UK (MPS); Cancer Research UK Clinical Trials Unit, Institute of Cancer Sciences, University of Glasgow, Glasgow, UK (JP, AH, JM); MRC Clinical Trials Unit at UCL, London, UK (RK); Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK (VE-P); Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia (NAA-T, BFM, SMW); Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK (RSH); NIHR Oxford Comprehensive Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK (DNC).

The authors have no disclosures. The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health, or the Wellcome Trust. We are grateful to the participants in the SCOT, COIN, and COIN-B trials who consented for the donation of samples for research, and to the investigators and pathologists who recruited patients and collected samples.

Author contributions: Study design: CP, JPC, IT, DNC; data collection: SB, CP, EJ, TI, RK, MPS, JP, AH, JM, MGS, EJ, HW, LG, TSM, RK, VE-P, NA-T, BFM, SMW, RSH; data analysis: VG, CP, JPC, IT, DNC; manuscript writing: DNC; manuscript approval: all authors.

References

1. Galon J, Costes A, Sanchez-Cabo F, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*. 2006;313(5795):1960–1964.
2. Pages F, Berger A, Camus M, et al. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med*. 2005;353(25):2654–2666.
3. Ghiringhelli F, Apetoh L, Tesniere A, et al. Activation of the NLRP3 inflammasome in dendritic cells induces IL-1beta-dependent adaptive immunity against tumors. *Nat Med*. 2009;15(10):1170–1178.
4. Tesniere A, Schlemmer F, Boige V, et al. Immunogenic death of colon cancer cells treated with oxaliplatin. *Oncogene*. 2010;29(4):482–491.
5. Vacchelli E, Ma Y, Baracco EE, et al. Chemotherapy-induced antitumor immunity requires formyl peptide receptor 1. *Science*. 2015;350(6263):972–978.

6. Vacchelli E, Enot DP, Pietrocola F, et al. Impact of pattern recognition receptors on the prognosis of breast cancer patients undergoing adjuvant chemotherapy. *Cancer Res*. 2016;76(11):3122–3126.
7. Apetoh L, Ghiringhelli F, Tesniere A, et al. Toll-like receptor 4-dependent contribution of the immune system to anticancer chemotherapy and radiotherapy. *Nat Med*. 2007;13(9):1050–1059.
8. Andre T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med*. 2004;350(23):2343–2351.
9. Takeuchi O, Akira S. Pattern recognition receptors and inflammation. *Cell*. 2010;140(6):805–820.
10. Netea MG, van der Meer JW. Immunodeficiency and genetic defects of pattern-recognition receptors. *N Engl J Med*. 2011;364(1):60–70.
11. Seifert R, Wenzel-Seifert K. The human formyl peptide receptor as model system for constitutively active G-protein-coupled receptors. *Life Sci*. 2003;73(18):2263–2280.
12. Ranjith-Kumar CT, Miller W, Sun J, et al. Effects of single nucleotide polymorphisms on Toll-like receptor 3 activity and expression in cultured cells. *J Biol Chem*. 2007;282(24):17696–17705.
13. Chen DN, Song CG, Yu KD, et al. A prospective evaluation of the association between a single nucleotide polymorphism rs3775291 in Toll-Like Receptor 3 and breast cancer relapse. *PLoS One*. 2015;10(7):e0133184.
14. Castro FA, Forsti A, Buch S, et al. TLR-3 polymorphism is an independent prognostic marker for stage II colorectal cancer. *Eur J Cancer*. 2011;47(8):1203–1210.
15. Senkus E, Kyriakides S, Ohno S, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2015;26(suppl 5):v8–v30.
16. Labianca R, Nordlinger B, Beretta GD, et al. Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2013;24(suppl 6):vi64–vi72.
17. Iveson TJ, Kerr RS, Saunders MP, et al. 3 versus 6 months of adjuvant oxaliplatin-fluoropyrimidine combination therapy for colorectal cancer (SCOT): an international, randomised, phase 3, non-inferiority trial. *Lancet Oncol*. 2018;19(4):562–578.
18. Adams RA, Meade AM, Seymour MT, et al. Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet Oncol*. 2011;12(7):642–653.
19. Wasan H, Meade AM, Adams R, et al. Intermittent chemotherapy plus either intermittent or continuous cetuximab for first-line treatment of patients with KRAS wild-type advanced colorectal cancer (COIN-B): a randomised phase 2 trial. *Lancet Oncol*. 2014;15(6):631–639.
20. Maughan TS, Adams RA, Smith CG, et al. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet*. 2011;377(9783):2103–2114.
21. Summers MG, Smith CG, Maughan TS, et al. BRAF and NRAS locus-specific variants have different outcomes on survival to colorectal cancer. *Clin Cancer Res*. 2017;23(11):2742–2749.
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–575.
23. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–909.
24. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011;9(2):179–181.
25. Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39(7):906–913.
26. Al-Tassan NA, Whiffin N, Hosking FJ, et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. *Sci Rep*. 2015;5:10442.
27. McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst*. 2005;97(16):1180–1184.
28. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60, 706 humans. *Nature*. 2016;536(7616):285–291.
29. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337(6090):64–69.
30. Walter K, Min JL, Huang J, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526(7571):82–90.
31. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst*. 2000;92(3):205–216.

European Journal of Cancer 124 (2020) 56–63

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Comprehensive analysis of colorectal cancer-risk loci and survival outcome: A prognostic role for *CDH1* variants☆

Matthew G. Summers [a], Timothy S. Maughan [b], Richard Kaplan [c], Philip J. Law [d], Richard S. Houlston [d], Valentina Escott-Price [e], Jeremy P. Cheadle [a,*]

[a] *Division of Cancer and Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK*
[b] *CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Roosevelt Drive, Oxford OX3 7DQ, UK*
[c] *MRC Clinical Trials Unit, Aviation House, 125 Kingsway, London, WC2B 6NH, UK*
[d] *Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SW7 3RP, UK*
[e] *Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff, CF24 4HQ, UK*

Received 28 May 2019; received in revised form 20 September 2019; accepted 28 September 2019

KEYWORDS
Colorectal cancer;
CDH1;
Survival;
Risk loci;
Prognostic biomarker

Abstract ***Purpose:*** Genome-wide association studies have identified common single nucleotide polymorphisms (SNPs) at 83 loci associated with colorectal cancer (CRC) risk in European populations. Because germline variation can also influence patient outcome, we studied the relationship between these SNPs and CRC survivorship.
Experimental design: For the 83 risk loci, 10 lead SNPs were directly genotyped, 72 were imputed and 1 was not genotyped nor imputed, in 1948 unrelated patients with advanced CRC from the clinical trials COIN and COIN−B (oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab). A Cox survival model was used for each variant, and variants classified by pathway, adjusting for known prognostic factors. We imposed a Bonferroni threshold of $P = 6.6 \times 10^{-4}$ for multiple testing. We carried out meta-analyses of published risk SNPs associated with survival.
Results: Univariate analysis identified six SNPs associated with overall survival (OS) ($P < 0.05$); however, only rs9939049 in *CDH1* remained significant beyond the Bonferroni threshold (Hazard Ratio [HR] 1.44, 95% Confidence Intervals [CI]: 1.21−1.71, $P = 5.0 \times 10^{-5}$). Fine mapping showed that rs12597188 was the most significant SNP at this locus and remained significant after adjustment for known prognostic factors beyond multiple

☆ This work was supported by Cancer Research Wales. The work of the Houlston laboratory was supported by Cancer Research UK (C1298/A8362). The COIN and COIN−B trials were funded by Cancer Research UK and an unrestricted educational grant from Merck-Serono.
* *Corresponding author*:
E-mail address: cheadlejp@cardiff.ac.uk (J.P. Cheadle).

https://doi.org/10.1016/j.ejca.2019.09.024

testing thresholds (HR 1.23, 95% CI: 1.13–1.34, $P = 1.9 \times 10^{-6}$). rs12597188 was also associated with poor response to therapy (OR 0.61, 95% CI: 0.42–0.87, $P = 6.6 \times 10^{-3}$). No combinations of SNPs within pathways were more significantly associated with survival compared with single variants alone, and no other risk SNPs were associated with survival in meta-analyses.

Conclusions: The CRC susceptibility SNP rs9939049 in *CDH1* influences patient survival and warrants further evaluation as a prognostic biomarker.

1. Introduction

Each year, over a million people are diagnosed with colorectal cancer (CRC) worldwide. Clinical stage, which combines depth of tumour invasion, nodal status and distant metastasis [1], is the only routinely used marker of survival. Other factors thought to influence prognosis include lifestyle [2,3], systemic inflammatory response to the tumour [4], the tumour immunologic microenvironment [5] and the patient's germline and the tumour's somatic genetic profile [6–9].

Around 6% of CRC is associated with Mendelian susceptibility caused by the inheritance of rare high-impact germline mutations [10] including those responsible for familial adenomatous polyposis (FAP; MIM 175100) [11], hereditary non-polyposis CRC (HNPCC; MIM 114500) [12] and MUTYH-associated polyposis (MAP; MIM 608456) [13]. Increasingly, it is being recognised that in addition to influencing CRC risk, germline variation plays a role in patient outcome with HNPCC and MAP-associated CRC typically being associated with better prognosis than those with sporadic CRC [14–16].

Genome-wide association studies (GWASs) have been successful in identifying single nucleotide polymorphisms (SNPs) robustly associated with an individual's risk of developing CRC. As well as influencing risk, studies have suggested that some of these alleles may affect patient survival [17–22]. However, most studies have not been performed in the context of a clinical trial but have been retrospective in design with the inherent biases from variation in patient management.

We have previously studied the relationship between SNP genotype and patient outcome for 14 of the GWAS risk loci by analysing patient data from two clinical trials—COIN and COIN–B [23]. Since this study, an additional 69 loci have been identified which influence CRC risk in European populations [24]. To gain a comprehensive understanding of the role of genetic variation on patient outcome, we assessed the prognostic effects of all known CRC risk SNPs in 1948 patients with advanced disease by further using COIN and COIN–B trial data.

2. Materials and methods

2.1. Samples

We prepared blood DNA samples from unrelated patients with metastatic or locally advanced colorectal adenocarcinoma from the MRC clinical trials COIN (NCT00182715) [25] and COIN–B (NCT00640081) [26]. All patients gave fully informed consent for bowel cancer research (approved by REC [04/MRE06/60]). COIN patients were randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy, continuous chemotherapy and cetuximab, or intermittent chemotherapy. COIN–B patients were randomised 1:1 to receive intermittent chemotherapy and cetuximab, or intermittent chemotherapy and continuous cetuximab.

2.2. Genotyping

As previously described [27], 2244 cases from COIN and COIN–B were genotyped using Affymetrix Axiom Arrays in accordance with the manufacturer's recommendations (Affymetrix, Santa Clara, CA 95051, USA). Individuals were excluded from analysis if they failed in one or more of the following thresholds: overall successfully genotyped SNPs <95% (n = 122), discordant sex information (n = 8), classified as out of bounds by Affymetrix (n = 30), duplication or cryptic relatedness (n = 4) and evidence of non-white European ancestry by PCA-based analysis (n = 130). After quality control, we had whole genome SNP genotyping and derived imputation data on 1950 patients, 2 of whom had no data on survival and were excluded (n = 1948). For the 83 CRC risk loci, 10 lead SNPs were directly genotyped, 72 were imputed and one (rs2732875) was on the X-chromosome which was not genotyped nor imputed. Six SNPs (rs77776598, rs2735940, rs6933790, rs704017, rs6055286 and rs1741640) had info scores <0.7 and were excluded.

2.3. Statistical analysis

We used a Cox survival model with overall survival (OS; time from trial randomisation to death) as the primary

measure. Univariate analyses were performed using *GenABEL* in R. Multivariate analyses were carried out using *survival* in R. The *coxph* function was used with prognostic covariates in COIN/COIN–B: sex (male vs. female: HR: 0.87, 95% CI; 0.78–0.97, $P = 9.7 \times 10^{-4}$), World Health Organization (WHO) performance status (HR: 1.42, 95% CI: 1.31–1.56, $P < 2.0 \times 10^{-16}$), resection status of the primary tumour (unresected/unresectable vs. local recurrence: HR: 1.29, 95% CI: 1.01–1.63, $P = 0.04$), white blood cell (WBC) count (HR: 1.03, 95% CI: 1.03–1.04, $P < 2.0 \times 10^{-16}$), platelet count (HR: 1.00, 95% CI: 1.00–1.00, $P < 2.0 \times 10^{-16}$), number of metastatic sites (HR: 1.21, 95% CI: 1.14–1.28, $P = 2.5 \times 10^{-10}$), site of distant metastasis (yes vs. no: liver, HR: 1.23, 95% CI: 1.09–1.39, $P = 8.8 \times 10^{-4}$; peritoneum, HR: 1.34, 95% CI: 1.16–1.54, $P = 6.4 \times 10^{-5}$; nodal, HR: 1.15, 95% CI: 1.04–1.28, $P = 7.5 \times 10^{-3}$; other metastases, HR: 1.31, 95% CI: 1.15–1.51, $P = 7.9 \times 10^{-5}$), *KRAS* status (mutant vs. wild type: HR: 1.46, 95% CI: 1.29–1.66, $P = 3.5 \times 10^{-9}$), *BRAF* status (mutant vs. wild type: HR: 2.29, 95% CI: 1.79–2.93, $P = 4.8 \times 10^{-11}$) and *NRAS* status (mutant vs. wild type: HR: 1.47, 95% CI: 1.08–1.99, $P = 0.01$), together with other factors in COIN/COIN–B (age at randomisation, cetuximab treatment, chemotherapy regimen, chemotherapy schedule, treatment arm and trial), none of which affected prognosis [25,28]. Response to treatment was defined as complete or partial response, and non-response was defined as stable or progressive disease at 12 weeks; analyses were performed with the *oddsratio* function from the *fmsb* package in R. We used Bonferroni correction to adjust for multiple testing with a significance threshold set at $P = 6.6 \times 10^{-4}$ (0.05/76 SNPs after exclusion of SNPs with poor imputation).

2.4. Meta-analyses

We collected published data for 6 CRC risk SNPs previously associated with survival, *albeit* at nominally significant levels ($P < 0.05$) (rs4939827 [17,19], rs961253 [18,22], rs6983267 [18,21], rs10795668 [17,20], rs4444235 [22] and rs4925386 [17,22]). These SNPs had been analysed in different cohorts: Colorectal Neoplasia Repository and North Central Cancer Treatment Group (NCCTG) [29]; Study on Colorectal Cancer in Scotland (SOCCS) [30]; Seattle Colon Cancer Family Registry (CCFR) [31]; Health Professionals Follow-up Study (HPFS), Nurses' Health Study (NHS), Physicians' Health Study (PHS), VITamins and Lifestyle Study (VITAL), Women's Health Initiative (WH1 and WH2) [17]; Nurses' Health Study (NHS), Health Professionals Follow-up Study (HPFS) [19]; and National Study of Colorectal Cancer Genetics (NSCCG) [22]. Meta-analyses were performed in R using the *meta* package. The metagen function was used to perform all analyses under a fixed effect model, or random effects model where there was significant heterogeneity. I^2 test and Cochran's Q tests were used for assessment of heterogeneity.

2.5. Bioinformatics

Linkage disequilibrium (LD) between SNPs was examined using the *ld* command in PLINK. Forty-nine SNPs were located within, or close to, genes (https://www.ncbi.nlm.nih.gov/snp). Thirteen SNPs were associated with expression quantitative trait loci (eQTL) (https://gtexportal.org/home/). Data from GeneCards (https://www.genecards.org) were used to assess whether ≥2 genes or eQTLs had roles in signalling pathways: 8 genes/eQTLs functioned in GPCR, 6 in TGF-Beta, 5 in ERK, 3 in Wnt, 3 in BMP, 3 in Hedgehog, 3 in PI3K-Akt, 3 in E-cadherin and 2 in Notch signalling. Combinations of SNPs within the same pathway were analysed for survival outcome by the log likelihood ratio test using the *coxph* and *anova* functions in R.

3. Results

We analysed blood DNA samples and survival data from 1948 unrelated patients with advanced CRC from the UK national trials COIN [25] and COIN–B [26] (Table 1). We found no evidence of heterogeneity in OS between patients when analysed by trial (COIN vs. COIN–B, $P = 0.33$), trial arm ($P = 0.49$), type of chemotherapy received (OxMdG/XELOX; $P = 0.46$),

Table 1
Clinicopathological data for patients in COIN and COIN–B.

	COIN	COIN–B
No. of cases with blood DNA	2078	196
No. with genotyping data after QC	1778	170
Total no. of deaths (% of cases)	1557 (75)	99 (51)
Median follow-up (SD)	2.4 (2.2)	2.0 (4.4)
% Female	34	42
Age at randomisation, N (%)		
<65 years	1203 (58)	115 (59)
65–69	422 (20)	35 (18)
70–74	318 (15)	31 (16)
75–79	124 (6)	10 (5)
≥80 years	9 (<1)	5 (3)
Missing	2 (<1)	0 (0)
Mean (SD)	62.0 (9.6)	61.7 (10.4)
Stage (%)		
1	0 (0)	0 (0)
2–3	0 (0)	0 (0)
4	2078 (100)	196 (100)
Unknown	0 (0)	0 (0)
Tumour site, N (%)		
Colon[a]	1103 (53)	124 (63)
Rectum[b]	951 (46)	71 (36)
Unknown	24 (1)	1 (1)

SD: standard deviation.
[a] Colon defined as caecum, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon and sigmoid colon.
[b] Rectum defined as rectosigmoid junction and rectum.

Table 2
Univariate analysis of CRC risk SNPs and overall survival.

Locus	SNP	Directly genotyped or imputed info score	Additive model			Recessive model		
			HR	95% CI	*P*	HR	95% CI	*P*
1p32.3	rs12143541	0.98	1.01	0.92–1.12	0.80	0.88	0.64–1.21	0.42
1p34.3	rs61776719	0.81	1.01	0.92–1.11	0.77	0.98	0.83–1.15	0.78
1p36.12	rs72647484	0.88	1.03	0.89–1.19	0.67	1.12	0.56–2.24	0.76
1q25.3	rs4546885	0.92	0.98	0.91–1.07	0.69	0.90	0.77–1.06	0.20
1q41	rs6658977	0.99	0.93	0.87–1.01	0.11	0.87	0.75–1.01	7.5×10^{-2}
2q11.2	rs11692435	0.85	1.01	0.85–1.19	0.92	2.48	1.23–4.97	0.01
2q33.1	rs11893063	0.92	1.05	0.96–1.14	0.28	1.10	0.96–1.26	0.19
2q33.1	rs7593422	0.98	1.01	0.94–1.09	0.77	1.00	0.88–1.14	0.99
2q35	rs13020391	0.95	0.96	0.89–1.04	0.36	0.89	0.76–1.04	0.13
3p21.1	rs9831861	1.00	0.94	0.87–1.01	0.11	0.86	0.75–1.00	4.6×10^{-2}
3p22.1	rs35470271	0.94	1.01	0.91–1.12	0.81	0.64	0.42–0.98	4.1×10^{-2}
3q13.2	rs12635946	0.97	0.98	0.91–1.06	0.60	0.97	0.83–1.13	0.71
3q26.2	rs35446936	0.85	0.98	0.88–1.08	0.65	1.03	0.74–1.42	0.87
4q24	rs17035289	DG	0.94	0.85–1.05	0.28	1.06	0.73–1.56	0.74
4q31.21	rs75686861	0.97	1.05	0.93–1.19	0.40	0.92	0.70–1.19	0.52
5p13.1	rs1445011	0.99	1.02	0.94–1.11	0.59	1.02	0.85–1.21	0.85
5q31.1	rs639933	0.80	1.01	0.91–1.12	0.87	1.01	0.81–1.26	0.95
6p12.1	rs62404966	0.97	1.05	0.96–1.14	0.30	1.21	0.98–1.48	7.2×10^{-2}
6p21.2	rs1321310	0.98	0.94	0.86–1.03	0.19	0.91	0.71–1.16	0.44
6p21.31	rs16878812	0.99	1.04	0.93–1.17	0.49	1.30	0.86–1.96	0.22
6p21.32	rs9271770	0.98	1.04	0.95–1.15	0.39	0.96	0.71–1.35	0.88
6p21.33	rs3131043	DG	1.03	0.95–1.10	0.50	1.01	0.88–1.15	0.91
6p24.1	rs2070699	DG	1.01	0.93–1.08	0.86	1.02	0.90–1.16	0.77
6q21	rs6928864	0.97	0.90	0.79–1.03	0.13	0.72	0.37–1.38	0.32
7p12.3	rs10951878	0.99	1.04	0.96–1.11	0.34	1.02	0.91–1.15	0.71
7p12.3	rs3801081	1.00	1.05	0.97–1.14	0.19	1.02	0.86–1.22	0.78
8q23.3	rs16892766	DG	1.23	1.08–1.39	1.3×10^{-3}	1.83	0.95–3.53	7.1×10^{-2}
8q24.21	rs6983267	DG	1.06	0.99–1.15	0.11	1.10	0.97–1.26	0.13
9p21.3	rs1412834	1.00	1.01	0.94–1.08	0.83	0.99	0.87–1.12	0.84
10p14	rs7894531	1.00	0.88	0.81–0.96	2.6×10^{-3}	0.77	0.63–0.93	8.4×10^{-3}
10q24.2	rs2193352	1.00	1.01	0.93–1.10	0.81	0.97	0.76–1.24	0.82
10q25.2	rs12255141	0.95	0.94	0.83–1.07	0.35	1.29	0.71–2.34	0.39
11p15.4	rs4450168	0.76	1.09	0.94–1.26	0.24	1.21	0.63–2.34	0.56
11q13.4	rs57796856	0.99	1.02	0.94–1.09	0.65	1.00	0.88–1.13	1.00
11q13.4	rs4944940	0.86	1.01	0.81–1.26	0.93	0.66	0.16–2.64	0.56
11q23.1	rs3087967	DG	1.05	0.98–1.14	0.18	1.12	0.94–1.32	0.21
12p13.31	rs10849438	0.89	1.00	0.88–1.14	0.97	1.37	0.79–2.36	0.26
12p13.32	rs12818766	0.96	0.96	0.87–1.06	0.43	1.08	0.79–1.46	0.64
12p13.32	rs3217810	0.73	0.98	0.84–1.14	0.75	0.73	0.35–1.54	0.41
12q13.13	rs11169572	0.99	1.05	0.97–1.13	0.24	1.12	0.98–1.28	0.10
12q13.3	rs7398375	0.73	0.99	0.87–1.11	0.91	0.83	0.61–1.14	0.26
12q24.12	rs597808	0.99	0.96	0.89–1.04	0.29	0.97	0.85–1.11	0.64
12q24.21	rs7315438	0.97	1.04	0.96–1.13	0.30	1.09	0.95–1.26	0.21
13q13.2	rs9537521	0.86	0.99	0.90–1.08	0.77	0.91	0.75–1.10	0.33
13q13.3	rs12427600	0.97	1.02	0.94–1.11	0.68	0.92	0.74–1.15	0.46
13q22.1	rs45597035	0.94	1.00	0.92–1.09	1.00	0.93	0.78–1.12	0.45
13q22.3	rs1330889	0.96	1.02	0.91–1.14	0.78	0.89	0.58–1.36	0.59
13q34	rs7993934	0.93	1.00	0.92–1.09	0.97	0.96	0.80–1.14	0.62
14q22.2	rs35107139	0.86	1.00	0.91–1.09	0.93	0.96	0.89–1.05	0.42
14q22.2	rs1570405	0.97	1.01	0.93–1.09	0.78	0.95	0.87–1.04	0.23
15q13.3	rs16969681	0.99	1.03	0.91–1.15	0.65	1.04	0.66–1.65	0.85
15q13.3	rs73376930	0.95	1.04	0.95–1.13	0.40	1.02	0.81–1.28	0.89
15q13.3	rs16959063	0.97	0.93	0.80–1.25	0.61	0.93	0.69–1.25	0.61
15q13.3	rs17816465	0.96	1.07	0.97–1.17	0.16	1.07	0.83–1.39	0.58
15q22.31	rs4776316	0.74	1.00	0.88–1.13	0.94	1.03	0.72–1.47	0.88
15q23	rs10152518	0.90	0.97	0.87–1.07	0.52	1.21	0.89–1.63	0.23
15q26.1	rs7495132	0.97	1.00	0.89–1.11	0.95	0.91	0.63–1.33	0.63
16q22.1	**rs9939049**	**1.00**	**1.12**	**1.03–1.21**	**8.1×10^{-3}**	**1.44**	**1.21–1.71**	**5.0×10^{-5}**
16q23.2	rs61336918	0.99	0.96	0.88–1.04	0.31	0.87	0.72–1.05	0.14
16q24.1	rs2696839	0.93	1.00	0.92–1.08	0.98	1.00	0.87–1.14	0.97
16q24.1	rs899244	0.96	1.02	0.93–1.12	0.68	0.92	0.70–1.20	0.53
17p12	rs1078643	0.85	0.94	0.84–1.05	0.30	1.45	1.03–2.03	3.2×10^{-2}

(*continued on next page*)

Table 2 (*continued*)

Locus	SNP	Directly genotyped or imputed info score	Additive model			Recessive model		
			HR	95% CI	*P*	HR	95% CI	*P*
17p13.3	rs73975588	0.97	1.05	0.93–1.19	0.43	0.96	0.54–1.69	0.88
18q21.1	rs7226855	DG	0.99	0.92–1.07	0.83	0.95	0.83–1.10	0.50
19p13.11	rs285245	0.98	0.95	0.79–1.14	0.57	NA	NA	NA
19q13.11	rs73039434	0.76	1.09	0.84–1.43	0.51	NA	NA	NA
19q13.2	rs9797885	0.91	1.00	0.92–1.09	0.94	1.05	0.95–1.16	0.35
19q13.33	rs12979278	0.92	0.98	0.91–1.06	0.65	1.02	0.95–1.10	0.51
20p12.3	rs961253	DG	0.99	0.92–1.07	0.84	0.97	0.84–1.13	0.71
20p12.3	rs6085661	0.99	1.00	0.93–1.08	0.90	1.12	0.97–1.28	0.11
20q13.12	rs2179593	0.98	0.96	0.88–1.05	0.37	0.89	0.72–1.11	0.31
20q13.13	rs6066825	DG	0.97	0.90–1.05	0.46	1.01	0.86–1.18	0.93
20q13.13	rs4811050	DG	1.01	0.92–1.11	0.86	1.00	0.77–1.31	0.98
20q13.13	rs1810502	0.83	1.07	0.98–1.17	0.15	1.07	0.90–1.26	0.44
20q13.13	rs6091213	0.91	1.03	0.94–1.12	0.54	1.01	0.82–1.25	0.92
20q13.33	rs3787089	0.78	1.03	0.93–1.14	0.54	1.03	0.82–1.31	0.79

HR: Hazard ratio, CI: Confidence interval, *P*: *P*-value, SNP: Single nucleotide polymorphism, NA: Not applicable. DG: Directly genotyped, CRC: Colorectal cancer.
Only rs9939049 at 16q22.1 (bold) was significant beyond the Bonferroni corrected threshold of $P = 6.6 \times 10^{-4}$.

or cetuximab use ($P = 0.24$), hence combined these groups for prognostic analyses. In total, 35% of patients were female with a mean age at randomisation of 63 years (range, 18–87 years, Table 1). We had over 70% power under an additive model to detect a HR of 1.18 for survival for SNPs with minor allele frequencies (MAFs) > 30% and a HR of 1.28 for SNPs with MAFs>10%.

For the 83 CRC risk loci, 10 lead SNPs were directly genotyped, 72 were imputed and 1 was on the X-chromosome which was not genotyped nor imputed. Univariate analyses identified six SNPs (rs9831861 at 3p21.1, rs35470271 at 3p22.1, rs16892766 at 8q23.3, rs7894531 at 10p14, rs9939049 at 16q22.1 and rs1078643 at 17p12) that were nominally associated with OS ($P < 0.05$) (Table 2). Only rs9939049 was significant beyond the Bonferroni corrected threshold (HR 1.44, 95% CI: 1.21–1.71, $P = 5.0 \times 10^{-5}$). rs9939049 lies within *CDH1* in LD with rs9929218 ($r^2 = 0.99$, D' = 0.99), which we have previously reported having a prognostic effect [23].

We consider whether other SNPs at 16q22.1 might be more significantly associated with survival and analysed all SNPs in LD with rs9939049 for which we had genetic data. rs12597188 (directly genotyped, $r^2 = 0.75$, D' = 0.99) was the most significantly associated SNP (recessive model: HR 1.48, 95% CI: 1.28–1.72, $P = 1.9 \times 10^{-7}$). We considered rs12597188 in multivariate analyses with known prognostic factors in COIN/COIN–B (sex, WHO performance status, resection status of the primary tumour, WBC and platelet count, number of metastatic sites, site of distant metastasis, and *KRAS*, *BRAF* and *NRAS* mutation status). rs12597188 remained significant beyond the Bonferroni corrected threshold (HR 1.23, 95% CI; 1.13–1.34, $P = 1.9 \times 10^{-6}$). Patients that were homozygous for the minor allele had a median decrease in life expectancy of 5 months compared to patients that were homozygous or heterozygous for the wild type allele.

We sought whether rs12597188 was associated with response to oxaliplatin-fluoropyrimidine chemotherapy after 12 weeks of treatment (likely to be correlated with survival, n = 1162 patients). Patients that were homozygous for the minor allele had significantly worse response (54/142 responded, 38.0%), as compared to patients that were heterozygous or homozygous wild type (512/1020 responded, 50.2%) (OR 0.61, 95% CI: 0.42–0.87, $P = 6.6 \times 10^{-3}$). This association was not seen in patients who also received cetuximab (51.0% versus 49.1%, n = 786) with significant heterogeneity between these groups ($I^2 = 75.2\%$, Cochran's Q test: $P = 0.04$).

We tested whether combinations of variants classified by pathway influenced survival. Eight SNPs lie within or near to genes that function in the GPCR signalling pathway, six in the TGF-Beta signalling pathway, five in the ERK signalling pathway, three in each of the Wnt, BMP, Hedgehog, PI3K-Akt and E-cadherin signalling pathways and two in the Notch signalling pathway (Table 3). No combinations of SNPs within specific pathways were more significantly associated with survival beyond the single most significant SNP in that pathway alone.

Six CRC risk SNPs (rs4939827, rs961253, rs6983267, rs10795668, rs4444235 and rs4925386) have previously been associated with survival [17–22], although none have been independently replicated [30,32,33]. We reviewed published survival data for these SNPs [17,19,22,29–31] and carried out meta-analysis with our data. No SNPs were associated with survival under fixed or random effects models (Fig. 1).

Table 3
Variants classified by signalling pathway.

Signalling pathway	SNPs
GPCR	rs2070699, rs4776316, rs3801081, rs9537521, rs35107139, rs6066825, rs73039434, rs62404966
TGF-Beta	rs62404966, rs12427600, rs4776316, rs35107139, rs7226855, rs73376930
ERK	rs4546885, rs62404966, rs7993934, rs35107139, rs9939049
Wnt	rs75686861, rs12427600, rs73376930
BMP	rs12427600, rs4776316, rs73376930
Hedgehog	rs75686861, rs12427600, rs73376930
PI3K-Akt	rs4546885, rs16878812, rs597808
E-cadherin	rs17816465, rs16959063, rs9939049
Notch	rs12427600, rs73376930

4. Discussion

Using an independent series of over 5000 cases with CRC, we have previously validated rs9929218 in *CDH1* as a prognostic biomarker [23]. We have extended these analyses *herein* and shown that rs12597188 is the most significantly associated SNP at this locus. Our data suggests that patients homozygous for the minor allele of rs12597188, equating to ~12% of patients, have worse survival, with a median decrease in life expectancy of 5 months (in the advanced disease setting). Another study has provided further support for *CDH1* variants having a genuine prognostic effect [20]. Our observations *herein* are limited to patients with stage 4 disease. It is

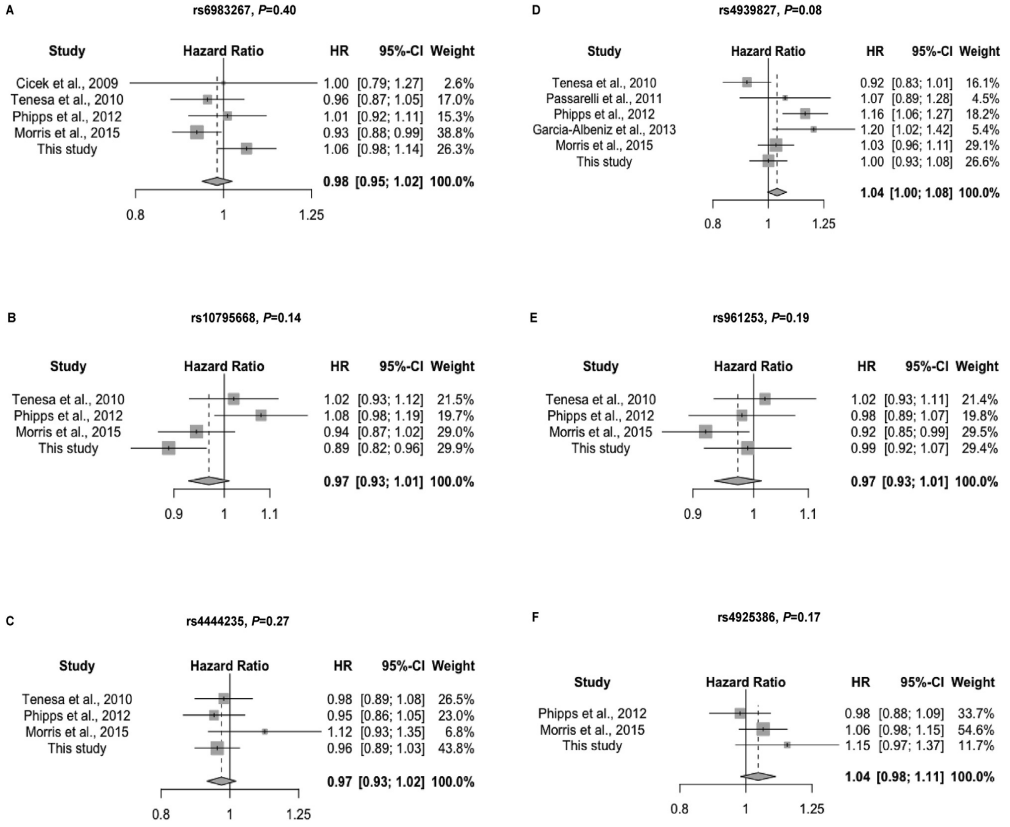

Fig. 1. Meta-analysis of six CRC risk SNPs previously associated with survival. Forest plots shown using a fixed effect model. rs10795668 and rs4939827 showed evidence of between-study heterogeneity ($I^2 = 72\%$, Cochran's Q $P = 0.01$ and $I^2 = 69\%$, Cochran's Q $P < 0.01$, respectively); however, neither were associated with survival when also considered under a random effects model ($P = 0.58$ and $P = 0.22$, respectively). Note - Results from Tenesa *et al.*, 2010 [30] and Morris *et al.*, 2015 [22] were not adjusted for prognostic factors; Cicek *et al.*, 2009 [29] adjusted for tumour characteristics at diagnosis, mismatch repair status, tumour site and stage; Passarelli *et al.*, 2011 [31] adjusted for age at diagnosis and race; Phipps *et al.*, 2012 [17] adjusted for age, and sex (VITamins and Lifestyle Study); Garcia-Albeniz *et al.*, 2013 [19] adjusted for age, race, sex, tumour stage, grade of differentiation, aspirin use, smoking status, alcohol consumption, consumption of meat, and, calcium and folate intake. The survival measure used by Phipps *et al.*, 2012 [17], Garcia-Albeniz *et al.*, 2013 [19], Tenesa *et al.*, 2010 [30] and Passarelli *et al.*, 2011 [31] was diagnosis to death with any cause mortality, Cicek *et al.*, 2009 [29] used overall survival or when censored at eight years, and Morris *et al.*, 2015 [22] used the date of recruitment to date of death or when censored at five years. Where appropriate, the inverse HR of those reported is shown to ensure the allele analysed for each study is consistent. HR: Hazard ratio. CI: Confidence Interval.

noteworthy that we have previously shown rs9929218 in *CDH1* was not associated with survival amongst patients with Stage 1–3 (premetastatic) disease (HR = 1.19, 95% CI: 0.93–1.52, $P = 0.18$) although there was no significant difference between the associations in patients with Stage 1–3 and Stage 4 disease ($P_{interaction} = 0.48$) [23]. Larger studies of premetastatic patients may help clarify the potential prognostic role of this biomarker in a population-based setting. It is also important to note that the effect sizes for *CDH1* variants are modest and will need to be combined with other germline and somatic prognostic factors to have any role in patient management; we are currently modelling potential combined effects in the advanced disease setting.

rs12597188, rs9939049 and rs9929218 are in strong LD with rs16260 [34] in the *CDH1* promoter, which downregulates *CDH1* expression [35]. *CDH1* encodes E-cadherin. Patients homozygous for the minor alleles of these variants would be expected to have reduced E-cadherin expression. E-cadherin functions as a transmembrane glycoprotein involved in intercellular adhesion, cell polarity and tissue morphology and regeneration [36]. Critically, its loss represents a defining feature of the epithelial to mesenchymal transition during metastasis. *CDH1* variants are therefore plausible prognostic biomarkers which influence this process.

Beyond *CDH1* variants, the next CRC risk loci most associated with survival was rs16892766 at 8q23.3. However, this variant was not significant after Bonferroni correction and was not significant in an independent cohort of >5000 patients with CRC [23]. Furthermore, our meta-analyses did not support a prognostic role for six other risk loci previously associated with survival. Given that our study was well powered to find variants with HRs>1.18, it is likely that no other low-penetrance CRC risk loci identified to date have clinically actionable effects on survival.

Role of the funding source

The study sponsors had no involvement in the study design, collection, analysis and interpretation of the data, the writing of the report nor the decision to submit this article for publication.

Declaration of competing interest

None declared.

Acknowledgements

The authors thank the patients and their families who participated and gave their consent for this research, and the investigators and pathologists throughout the UK who submitted samples for assessment. COIN and COIN–B were coordinated by the Medical Research Council Clinical Trials Unit and conducted with the support of the National Institute of Health Research Cancer Research Network.

References

[1] Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer 2009;9:489–99.
[2] Haydon AM, MacInnis RJ, English DR, Giles GG. Effect of physical activity and body size on survival after diagnosis with colorectal cancer. Gut 2006;55:62–7.
[3] Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D, et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ 2007;335:1134.
[4] Leitch EF, Chakrabarti M, Crozier JE, McKee RF, Anderson JH, Horgan PG, et al. Comparison of the prognostic value of selected markers of the systemic inflammatory response in patients with colorectal cancer. Br J Canc 2007;97:1266–70.
[5] Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Largorce-Pagès C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 2006;313:1960–4.
[6] Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol 2005;23:609–18.
[7] Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut 2008;57:941–50.
[8] Lochhead P, Kuchiba A, Imamura Y, Liao X, Yamauchi M, Nishihara R, et al. Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. J Natl Cancer Inst 2013;105:1151–6.
[9] Eklöf V, Wikberg ML, Edin S, Dahlin AM, Jonsson BA, Öberg Å, et al. The prognostic role of KRAS, BRAF, PIK3CA and PTEN in colorectal cancer. Br J Canc 2013;108:2153–63.
[10] Aaltonen L, Johns L, Jarvinen H, Mecklin JP, Houlston R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. Clin Cancer Res 2007;13:356–61.
[11] Fearnhead NS, Britton MP, Bodmer WF. The ABC of APC. Hum Mol Genet 2001;10:721–33.
[12] Peltomäki P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum Mol Genet 2001;10:735–40.
[13] Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, et al. Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. Nat Genet 2002;30:227–32.
[14] Watson P, Lin KM, Rodriguez-Bigas MA, Smyrk T, Lemon S, Shashidharan M, et al. Colorectal carcinoma survival among hereditary nonpolyposis colorectal carcinoma family members. Cancer 1998;83:259–66.
[15] Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 2006;354:2751–63.
[16] Nielsen M, van Steenbergen LN, Jones N, Vogt S, Vasen HF, Morreau H, et al. Survival of MUTYH-associated polyposis patients with colorectal cancer and matched control colorectal cancer patients. J Natl Cancer Inst 2010;102:1724–30.
[17] Phipps AI, Newcomb PA, Garcia-Albeniz X, Hutter CM, White E, Fuchs CS, et al. Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. Gastroenterology 2012;143:51–4.

[18] Dai J, Gu J, Huang M, Eng C, Kopetz ES, Ellis LM, et al. GWAS-identified colorectal cancer susceptibility loci associated with clinical outcomes. Carcinogenesis 2012;33:1327–31.
[19] Garcia-Albeniz X, Nan H, Valeri L, Morikawa T, Kuchiba A, Phipps AI, et al. Phenotypic and tumor molecular characterization of colorectal cancer in relation to a susceptibility SMAD7 variant associated with survival. Carcinogenesis 2013;34:292–8.
[20] Abulí A, Lozano JJ, Rodríguez-Soler M, Jover R, Bessa X, Munoz J, et al. Genetic susceptibility variants associated with colorectal cancer prognosis. Carcinogenesis 2013;34:2286–91.
[21] Takatsuno Y, Mimori K, Yamamoto K, Sato T, Niida A, Inoue H, et al. The rs6983267 SNP is associated with MYC transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. Ann Surg Oncol 2013;20:1395–402.
[22] Morris EJ, Penegar S, Whiffin N, Broderick P, Bishop DT, Northwood E, et al. A retrospective observational study of the relationship between single nucleotide polymorphisms associated with the risk of developing colorectal cancer and survival. PLoS One 2015;10. e0117816.
[23] Smith CG, Fisher D, Harris R, Maughan TS, Phipps AI, Richman SD, et al. Analyses of 7,635 patients with colorectal cancer using independent training and validation cohorts show that rs9929218 in CDH1 is a prognostic marker of survival. Clin Cancer Res 2015;21:3453–61.
[24] Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun 2019;10:2154.
[25] Maughan TS, Adams RA, Smith CG, Meade AM, Seymour MT, Wilson RH, et al. The addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. Lancet 2011;377:2103–14.
[26] Wasan H, Meade AM, Adams R, Wilson R, Pugh C, Fisher D, et al. Intermittent chemotherapy plus either intermittent or continuous cetuximab for first-line treatment of patients with KRAS wild-type advanced colorectal cancer (COIN-B): a randomised phase 2 trial. Lancet Oncol 2014;15:631–9.
[27] Al-Tassan NA, Whiffin N, Hosking FJ, Palles C, Farrington SM, Dobbins SE, et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. Sci Rep 2015;5:10442.
[28] Smith CG, Fisher D, Claes B, Maughan TS, Idziaszczyk S, Peuteman G, et al. Somatic profiling of the epidermal growth factor receptor pathway in tumors from patients with advanced colorectal cancer treated with chemotherapy ± cetuximab. Clin Cancer Res 2013;19:4104–13.
[29] Cicek MS, Slager SL, Achenbach SJ, French AJ, Blair HE, Fink SR, et al. Functional and clinical significance of variants localized to 8q24 in colon cancer. Cancer Epidemiol Biomark Prev 2009;18:2492–500.
[30] Tenesa A, Theodoratou E, Din FV, Farrington SM, Cetnarskyj R, Barneston RA, et al. Ten common genetic variants associated with colorectal cancer risk are not associated with survival after diagnosis. Clin Cancer Res 2010;16:3754–9.
[31] Passarelli MN, Coghill AE, Hutter CM, Zheng Y, Makar KW, Potter JD, et al. Common colorectal cancer risk variants in SMAD7 are associated with survival among prediagnostic nonsteroidal anti-inflammatory drug users: a population-based study of postmenopausal women. Genes Chromosomes Cancer 2011;50:875–86.
[32] Hoskins JM, Ong PS, Keku TO, Galanko JA, Martin CF, Coleman CA, et al. Association of eleven common, low-penetrance colorectal cancer susceptibility genetic variants at six risk loci with clinical outcome. PLoS One 2012;7. e41954.
[33] Sanoff HK, Renfro LA, Poonnen P, Ambadwar P, Sargent DJ, Goldberg RM, et al. Germline variation in colorectal risk loci does not influence treatment effect or survival in metastatic colorectal cancer. PLoS One 2014;9. e94727.
[34] Pittman AM, Twiss P, Broderick P, Lubbe S, Chandler I, Penegar S, et al. The CDH1-160C>A polymorphism is a risk factor for colorectal cancer. Int J Cancer 2009;125:1622–5.
[35] Li LC, Chui RM, Sasaki M, Nakajima K, Perinchery G, Au HC, et al. A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities. Cancer Res 2000;60:873–6.
[36] Takeichi M. Cadherin cell adhesion receptors as a morphogenetic regulator. Science 1991;251:1451–5.