Identifying genetic biomarkers of response to treatment for advanced colorectal cancer

A thesis submitted in candidature for the degree of
Doctor of Philosophy (PhD)

Victoria Gray

2021

Division of Cancer and Genetics
School of Medicine
Cardiff University

Supervisors

Prof. Jeremy Cheadle
Prof. Valentina Escott-Price

Contents

Abstract

Background

Although there are tumour-based biomarkers of response to chemotherapy and cetuximab, there is a lack of germline predictive biomarkers. I sought such biomarkers by analysing patients from the COIN and COIN-B clinical trials who received oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab for advanced colorectal cancer (aCRC).

Patients and Methods

2,244 blood DNA samples were genotyped on whole genome arrays and imputed for 5 million common genetic variants (SNPs); 1,649 patients had data on response at 12 weeks and 1,948 had data on overall survival (OS). SNPs in pattern recognition proteins (PRPs) were analysed by Cox regression. Univariate and multivariate genome-wide association studies (GWASs) for response to oxaliplatin and fluoropyrimidine-based chemotherapy were performed. To identify predictive biomarkers for cetuximab, I performed exploratory factor analyses (including *RAS [KRAS and NRAS]* mutational status and type of chemotherapy) and univariate and multivariate GWASs in 319 patients with *RAS* wild-type CRCs.

Results

Loss of function SNPs in PRP genes were not associated with benefit from oxaliplatin-based chemotherapy. Genome wide analyses identified five loci suggestive of association ($P<1x10^{-5}$) with response to chemotherapy and SNPs at 10p15.3 (*WDR37* and an eQTL for *IDI1*) influenced OS (lead SNP rs2086382, HR=0.77, 95% CI=0.65-0.92, $P=3.0x10^{-3}$). *RAS* mutation status was predictive for response to cetuximab ($P_{Interaction}<0.01$); 71% of patients with *RAS* wild-type CRCs responded to chemotherapy plus cetuximab *versus* 61% without cetuximab (OR=1.61, 95% CI=1.19–2.19, $P<0.01$). Although not genome-wide significant, rs12054810 (eQTL for *ISCO1*, $P=3.0x10^{-6}$), rs142144203 (*RNLS*, $P=9.7x10^{-6}$), rs73200904 (*PCDH9*, $P=5.7x10^{-6}$) and rs131850 ($P=7.7x10^{-6}$) genotypes showed evidence for an association with response to cetuximab.

Conclusions

Cetuximab improves response in patients with *RAS* wild-type aCRC and I have identified potential biomarkers and their pathways to optimise therapy for both cetuximab and oxaliplatin-based chemotherapy.

Acknowledgements

"None of us got to where we are alone. Whether the assistance we received was obvious or subtle" – Harvey Mackay

I would like to thank the following:

My supervisors Professor Jeremy Cheadle and Professor Valentina Escott-Price for the guidance, supervision and encouragement throughout my PhD.

The patients and all involved in the investigation of the COIN and COIN-B clinical trials, without whom this project would not have been possible.

Cardiff University School of Medicine for funding this project. Dr. Amanda Tonks for her support and kindness. The support staff in the Postgraduate Office for their continued assistance.

Dr. David Church for his input, advice and collaboration as co-author for the published work in this thesis. David Fisher for his help and advice regarding the statistical analyses. Dr. Emily Baker and Dr. Ganna Leonenko's advice about methods and data interpretation.

Pamela Smith, Chris Wills, Katie Watts, Dr. Hannah West, Dr. Matthew Summers, Dr. Elena Meususer, Dr. Alex Georgiades, Dr. Zoe Hudson, Dr. Magda Meissner, Dr. Gareth Marlow and Dr. Martin McClarty for support, advice and chats over coffee.

Grace Balfour-Harle for her excellent proofreading skills and unwavering friendship. My other amazing friends – Hannah, Jess, Giuliana, Roisin, Abbi, Elle and Olivia (just to name a few) – who have cheered me on from across the world. My brother, Matthew, and my extended family for their interest and enthusiasm for all I do.

And finally my parents – the reason I am who I am today. They have shown so much faith in me, provided endless encouragement and cheered me up with terrible jokes.

Abbreviations

Table I. Abbreviations for amino acids

Abbreviation	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 II. Other Abbreviations

Abbreviation	Description
-	Deletion
%	Percent
↑	Increased
↓	Decreased
λ	Genomic inflation factor
5-FU	5-fluorouracil
95% CI	95% confidence interval
aCRC	Advanced colorectal cancer
AJCC	American Joint Committee on Cancer
ALFA	Allelic frequency aggregator
ALKP	Alkaline phosphatase
ALT	Alternative allele
ANOVA	Analysis of variance
APC	*Adenomatous polyposis coli*
ARCCA	Advanced Research Computing at Cardiff
ATM	*Ataxia Telangiectasis Mutated*
BMI	Body Mass Index
BP	Base position
BRAF	*v-Raf murine sarcoma viral oncogene homolog B*
c.	Coding
CD/CV	Common disease, common variant
Chr	Chromosome
CI	Confidence interval
CIMP	CpG Island Methylator Phenotype
CIN	Chromosomal instability
COIN	COntinuous versus INtermittent
CRAN	Comprehensive R Archive Network
CRC	Colorectal cancer
CSS	Cancer-specific survival
DAMP	Damage-associated molecular pattern
DC	Dendritic cells
DDX52	*DExD/H-Box Helicase 52*
DDX58	*DExD/H-Box Helicase 58*
DFS	Disease-free survival
DNA	Deoxyribose nucleic acid
dTMP	Deoxythmidine monophosphate
dUMP	Deoxyurudine monophosphate
ECM	Extracellular matrix
EGF	Epidermal Growth Factor
EGFR	*Epidermal Growth Factor Receptor*
EMA	European Medicines Agency
ENCODE	The Encyclopedia of DNA Elements
eQTL	Expressive quantitative trait loci
ERCC1	*Excision Repair Cross-Complementation Group 1*
ERCC5	*Excision Repair Cross-Complementation Group 5*

Abbreviations

FA	Folinic acid
FAP	Familial adenomatous polyposis
FDA	US Food and Drug Administration
FDR	False discovery rate
FFPE	Formalin fixed paraffin embedded
FFS	Failure-free survival
FOLFIRI	Folinic acid, fluorouracil, irinotecan (chemotherapy regimen)
FOLFOX	Folinic acid, fluorouracil, oxaliplatin (chemotherapy regimen)
FP	Fluoropyrimidine
FPR1	*Formyl Peptide Receptor 1*
FPR3	*Formyl Peptide Receptor 3*
GAS	Growth-arrest specific
GO	Gene ontology
GOF	Gain of function
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study
HDI	Human Development Index
HNPCC	Hereditary non-polyposis colon cancer (also known as Lynch syndrome)
HPC	High-performance cluster
HR	Hazard ratio
HWE	Hardy-Weinberg Equilibrium
ICR	Institute of Cancer Research
IDE	Integrated development environment
IDI1	*Isopentenyl-Diphosphate Delta Isomerase 1*
IFIH1	*Interferon Induced With Helicase C Domain 1*
IFL	Irinotecan, fluorouracil and leucovorin
Indels	Insertions and deletions
ISOC1	*Isochorismatase Domain Containing 1*
ITT	Intention to treat
KRAS	*Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*
LD	Linkage disequilibrium
LOF	Loss of function
LV	Leucovorin
mAb	Monoclonal antibody
MAF	Minor allele frequency
MAP	MUTYH-associated polyposis
mCRC	Metastatic colorectal cancer
MFS	Metastasis-free survival
MMP2	*Matrix metalloproteinase-2*
MMR	Mismatch repair
MRC	Medical Research Council
mRNA	Messenger RNA
MSI	Microsatellite instability
MSS	Microsatellite stable
Mut	Mutant
MWAS	Methylome-wide association study
N	Number

N/A	Not applicable
NCBI	National Centre for Biotechnology Information
NES	Normalised effect sizes
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NK	Natural killer (cell)
n/k	Not known
NLR	NOD-I-like recptors
no cet	Patients who did not receive cetuximab
NOD2	*Nucleotide-binding oligomerization domain-containing protein 2*
NRAS	*Neuroblastoma RAS viral oncogene homolog*
NSD	No significant difference
OR	Odds ratio
ORR	Overall recurrence rate
OS	Overall survival
OxMdG	Oxaliplatin modified de Gramont (chemotherapy regimen)
p	Short arm of chromosome
P	*P*-value
p.	Protein
PAMP	Pathogen-associated molecular pattern
PCA	Principle Component Analyses
PCDH9	*Protocadherin 9*
PFS	Progression-free survival
PRP	Pattern recognition protein
PRR	Pattern recognition receptor
PRS	Polygenic risk score
PTMA	*Human prothymosin-α*
Q	*Q*-value
q	Long arm of chromosome
Q-Q	Quantile-quantile
QC	Quality control
RAS	*KRAS* and *NRAS*
RCT	Randomly controlled trial
RECIST	Response Evaluation Criteria In Solid Tumours
REF	Reference allele
REMARK	REporting recommendations for tumour MARKer prognostic studies
RFS	Recurrence-free survival
RR	Risk ratio
SCOT	3 vs 6 months of adjuvant oxaliplatin fluoropyrimidine combination therapy for CRC
SD	Standard deviation
SDM	Shared decision making
SE	Standard error
SFFS	Strategy-failure-free survival
SMAD4	*SMAD Family Member 4*
SNP	Single nucleotide polymorphism
SOX	S-1 and oxaliplatin

STR	Short tandem repeat
Surv	Survival
T	Tyrosine
TAM	Tumour-associated macrophages
TF	Transcription factor
TFBS	Transcription factor binding sites
TGF- α	Transforming growth factor α
TLR2	*Toll-like receptor 2*
TLR3	*Toll-like receptor 3*
TLR4	*Toll-like receptor 4*
TLR9	*Toll-like receptor 9*
TO	Treatment outcome
TP53	*Tumour Protein P53*
tRNA	Transfer RNA
TS	Thymidylate Synthase
TSG	Tumour surpressor gene
TWAS	Transcriptome wide association study
UC	Ulcerative colitis
UICC	Union for International Cancer Control
UK	United Kingdom
USA	United States of America
VEGF	*Vascular endothelial growth factor*
WBC	White blood cell
WDR37	*WD Repeat Domain 37*
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organisation
wt	Wild-type
XELIRI	Irinotecan and capecitabine
XELOX	Oxaliplatin and capecitabine (chemotherapy regime)

List of figures

Figures

List of tables

Tables

Publications

Publications as a direct result of the works in this thesis:

Gray V, Briggs S, Palles C, Jaeger E, Iveson T, Kerr R, Saunders MP, Paul J, Harkin A, McQueen J, Summers MG, Johnstone E, Wang H, Gatcombe L, Maughan TS, Kaplan R, Escott-Price V, Al-Tassan NA, Meyer BF, Wakil SM, Houlston RS, Cheadle JP, Tomlinson I, Church DN. Pattern Recognition Receptor Polymorphisms as Predictors of Oxaliplatin Benefit in Colorectal Cancer. J Natl Cancer Inst. 2019 Aug 1;111(8):828-836.

Publications as a result of additional work I have been involved in during the course of this PhD project:

Alessi AM, **Gray V**, Farquharson FM, Flores-López A, Shaw S, Stead D, Wegmann U, Shearman C, Gasson M, Collie-Duguid ESR, Flint HJ, Louis P. β-Glucan is a major growth substrate for human gut bacteria related to Coprococcus eutactus. Environ Microbiol. 2020 Jun;22(6):2150-2164.

Watts K, Wills C, Madi A, Palles C, Maughan TS, Kaplan R, Al-Tasaan NA, Kerr R, Kerr, D, **Gray V**, West H, Houlston, RS, Escott-Price V. & Cheadle, JP 2021. Genome-wide association studies of toxicity to oxaliplatin and fluoropyrimidine chemotherapy with or without cetuximab in 1800 patients with advanced colorectal cancer. Int J Cancer, 149, 1713-1722.

1. Chapter 1: Introduction

1.1 Colorectal cancer

1.1.1 Incidence and mortality

Colorectal cancer (CRC) -– cancer of large bowel, including the colon and rectum – is the second most common cause of cancer death in the United Kingdom (UK) and is the fourth most common cause of cancer-related death worldwide (Office for National Statistics, 2020, Rawla et al., 2019). Based on 2020 data, approximately 1.9 million new cases of CRC are diagnosed annually making it the third most commonly diagnosed cancer worldwide, accounting for >10% of all new cancer diagnoses (Sung et al., 2021). An estimated 61% of CRC cases originate in the colon with the remaining 39% cases forming in the rectum (Rawla et al., 2019). It has been predicted that by 2030, the global burden of CRC could double from an estimated 1.1 million to >2.2 million newly diagnosed cases and >1 million CRC related deaths (Arnold et al., 2017). This increase in incidence and mortality has been attributed to a number of factors including smoking, diet and sedentary lifestyle – which can contribute to increased rates of obesity – and an aging population (Kuipers et al., 2015). These factors are associated with relative developmental status of countries and high rates of CRC incidence, which is supported by Europe, North America, Australia and Eastern Asia being areas with the highest relative rates of new diagnoses (Rawla et al., 2019). Additionally, as Human Development Index (HDI, a measure of development including life expectancy, education and income indicators) scores rise, CRC incidence rises at a similar rate, with evidence that this may be a causal relationship (Desai, 1991, Rafiemanesh et al., 2016).

Of all cases diagnosed in Europe, it is estimated that 50% of CRC patients will develop metastases, with around half of these patients (20–25% of all CRC cases) presenting with metastases at diagnosis (Haggar and Boushey, 2009, Riihimäki et al., 2016). This is likely because early diagnosis of CRC can be a challenge. Many CRC cases are asymptomatic until later stages and symptoms that present early such as anaemia, constipation and fatigue are common in other non-cancerous disorders that are more

common than CRC in adults (Dekker et al., 2019, Fletcher, 2009). Whilst some developed countries (such as the UK) are delivering screening programs to combat this delay in diagnosis, uptake for these can be poor. Just over 60% of those invited for at home screening by the National Health Service (NHS) in England and Wales complete the process (Public Health England, 2020, Public Health Wales, 2020). This pattern of late diagnosis is a key contributor to the poor prognosis of CRC. Recent statistics available for England show a 5-year survival rate for all stages of CRC of 58.4% but this drops to 10.3% in patients diagnosed with metastatic CRC (mCRC) (Office for National Statistics, 2019). Despite these statistics, the outcomes for patients treated for mCRC have improved from six months with best supportive care alone to over 30 months (Scheithauer et al., 1993, Van Cutsem et al., 2016). This is likely due to increased surgical resection, more targeted approaches to treatment and earlier diagnosis of CRC (Van Cutsem et al., 2016).

1.1.2 CRC staging

CRC prognosis is significantly associated with tumour stage at diagnosis. Therefore, accurate methods of CRC staging are vital for understanding prognosis and determining appropriate treatments. Most classification of CRC tumours is performed using the American Joint Committee on Cancer (AJCC) staging manual, currently on the 8th edition (Amin et al., 2017). This has been, in conjunction with the AJCC's partner, Union Internationale Control le Cancer (UICC), successfully deployed worldwide for over 20 years and is widely considered the 'gold standard' for tumour staging (Brierley et al., 2017, Yarbro et al., 1999, Benson et al., 2017). This method, also called TMN staging, classifies tumours based on invasion depth (T stage), involvement of lymph nodes (N) and presence of metastatic sites (M, Table 1.1). These are combined into an overall CRC stage definition which is used to help determine appropriate therapeutic actions (Brenner et al., 2014). TNM staging has largely been adopted due to its categorical nature which is simple to follow and its link to patient outcomes such as overall survival (OS). This has resulted in it being the principal classification for solid tumours such as CRC (Kattan et al., 2016).

Table 1.1. Pathologic staging of colorectal carcinoma and corresponding descriptions

Stage		TNM Staging			Description
		Tumour size (T)	Lymph nodes (N)	Metastasis (M)	
0		Tis	N0	M0	Cancer confined to mucosa
I		T1	N0	M0	Tumour infiltrates *submucosa*
		T2	N0	M0	Tumour infiltrates *muscularis propria*
II	IIA	T3	N0	M0	Tumour infiltrates *subserosa* and beyond
	IIB	T4a	N0	M0	Tumour infiltrates *serosa*
	IIB	T4b	N0	M0	Tumour infiltrates neighbouring tissues
III	IIIA	T1-T2	N1	M0	Tumour infiltrates up to *muscularis propria*. Cancer in 1–3 lymph nodes
		T1	N2a	M0	Tumour infiltrates *submucosa.* Cancer in 4–6 regional lymph nodes
	IIIB	T3-T4a	N1	M0	Tumour infiltrates up to *serosa*. Cancer in 1–3 regional lymph nodes
		T2-T3	N2a	M0	Tumour up to *subserosa* and beyond. Cancer in 4–6 regional lymph nodes
		T1-T2	N2b	M0	Tumour infiltrates up to *muscularis propria*. Cancer in ≥7 regional lymph nodes
	IIIC	T4a	N2a	M0	Tumour infiltrates *serosa*. Cancer in 4–6 regional lymph nodes
		T3-T4a	N2b	M0	Tumour infiltrates up to *serosa*. Cancer in ≥7 regional lymph nodes
		T4b	N1-N2	M0	Tumour infiltrates neighbouring tissues. Cancer in regional lymph nodes
IV	IVA	Any	Any	M1a	Metastasis to one distant organ or distant lymph nodes
	IVB	Any	Any	M1b	Metastasis to more than one distant organ or peritoneal metastasis

Based on Overall Union Internationale Control le Cancer stage classification of colorectal cancers. Adapted from Brenner et al. (2014).

Work in this thesis primarily focuses on patients with stage IV CRC, sometimes referred to as mCRC or advanced CRC (aCRC). Patients who reach stage IV are largely treated with palliative or end of life treatments as most patients are beyond the reach of curative therapy (Wasserberg and Kaufman, 2007, Van Cutsem et al., 2014).

1.1.3 Colorectal cancer formation and outcomes

Many cancers, including CRC, are complex diseases influenced by both genetic and lifestyle factors (Anand et al., 2008). It has argued that genetics may account for between 5–20% of all worldwide cancer cases, with the remaining cases the result of environmental factors (Anand et al., 2008, Parsa, 2012, Pomerantz and Freedman, 2011).

Some cancer types have a single risk factor that accounts for the majority of cancer cases, such as the role of tobacco in the formation of lung cancer (de Groot et al., 2018, Peto et al., 1992). However, this is not the case with the formation of CRC (Brenner et al., 2014). Contributory factors to colorectal tumorigenesis include dietary and other lifestyle factors, germline genetic variation and somatic mutations (Kuipers et al., 2015, Fearon, 2011, Bogaert and Prenen, 2014).

1.1.3.1 Demographic and lifestyle factors

Specific environmental factors – including behavioural, demographic and lifestyle factors – have been shown to contribute to CRC development (Section 1.1.1, (Rawla et al., 2019).

Other demographic factors have been shown to influence CRC incidence and outcomes. These include sex, with multiple studies showing increased cancer incidence and deaths in males; and age, diagnoses of sporadic CRC in individuals <40 years of age is uncommon and over 70s have the highest rates of incidence (White et al., 2018, Douaiher et al., 2017, Sung et al., 2021, Steele et al., 2014, Millan et al., 2015). There has also been evidence in developed countries such as the United States of America (USA), that the incidence and mortality of CRC is greater in

individuals of African descent compared to white individuals from the same areas (Augustus and Ellis, 2018). Lower socioeconomic status has also been shown to increase incidence and mortality across multiple cancer types, independent of comorbidity (Augustus and Ellis, 2018, Siegel et al., 2008, Doubeni et al., 2012).

Some of these socioeconomic factors can also contribute behaviours that influence cancer incidence, stage at diagnosis and mortality. A key behaviour shown to directly correlate with poor prognosis is early screening attendance (Section 1.1.1), with deprivation having a significant impact on screening uptake (Quyn et al., 2018, von Wagner et al., 2009). Lack of adherence to recommended screening interventions also correlates with other behaviours found to increase the risk of CRC development (Coups et al., 2007).

A key lifestyle factor associated with behaviour is increased body mass index (BMI) which can be associated with lack of education over diet and decreased physical activity, also potentially increasing CRC risk (Shaukat et al., 2017, Torres Stone et al., 2017, Shaw et al., 2018). Other socioeconomic-associated dietary components linked to CRC risk include consumption of highly processed foods and drinks and alcohol (Fiolet et al., 2018, Schwingshackl et al., 2018, McNabb et al., 2020).

New research indicates the influence of non-genetic factors vary by primary tumour site. For example, Demb et al. (2019) has shown that males have increased risk of developing CRC but with a significant difference based on primary tumour site (Rectal Odds Ratio (OR)=2.85, 95% CI=2.25-3.58, $P<0.05$; Distal Colon OR=1.84, 95 CI 1.50-2.24, $P<0.05$).

1.1.3.2 Biological and genetic factors

Biological factors have also been shown to increase CRC risk. These can include development of other diseases including familial adenomatous polyposis (FAP), ulcerative colitis (UC) and inflammatory bowel disease (IBD) (Jang et al., 1997, Yashiro, 2014, Kim and Chang, 2014). A family history CRC increases the risk of

disease development, partly because of a genetic element of CRC (Lynch and de la Chapelle, 2003, Munteanu and Mastalier, 2014).

Investigation of both sporadic CRC and inherited CRC syndrome development has resulted in greater understanding of the genetic and molecular mechanisms underpinning these diseases (Fearon, 2011). Some of this has been described by Hanahan and Weinberg (2011) and detailed in Figure 1.1.

The three most common inherited CRC syndromes are FAP, MUTYH-associated polyposis (MAP) and Hereditary Non-Polyposis Colorectal Cancer (HNPCC, also known as Lynch syndrome) (Galiatsatos and Foulkes, 2006, Poulsen and Bisgaard, 2008, Chung and Rustgi, 2003). It is estimated that <10% of all CRC cases are as a result of inherited syndromes such as these (Kastrinos and Syngal, 2011, Schlussel et al., 2014, Snyder and Hampel, 2019).

There are two types of genetic mutations that initiate sporadic tumorigenesis across many cancer types (including CRC) – loss of function (LOF) of tumour-suppressor genes (TSGs) and gain of function (GOF) of oncogenes (Armaghany et al., 2012, Slattery et al., 2017). Examples of genes found in CRC are shown in Table 1.2.

TSGs can be classified by the role they play in the regulation of key cellular functions including cell cycle progression, proliferation, growth and mechanisms of DNA repair (Joyce et al., 2021, Knudson, 1993). TSGs can be separated into three classes – caretakers, gatekeepers and landscapers (Kinzler and Vogelstein, 1997, Michor et al., 2004). Caretaker genes are responsible for maintaining genomic integrity of cells through DNA repair processes and cell-cycle checkpoints (Levitt and Hickson, 2002). Gatekeeper genes regulate cell division, proliferation and death, ensuring the balance between these processes is appropriate when cellular damage occurs (Kinzler and Vogelstein, 1997, Frank, 2003, Kinzler and Vogelstein, 1996). Landscaper genes do not directly influence cellular growth and their role in cellular regulation involves microenvironmental components including the tumour extracellular matrix (ECM) (Knudson, 1971, Knudson, 1996). In FAP, one of these alleles is a germline mutation (Leoz et al., 2015). In cases of sporadic CRC, both of the mutations are somatic.

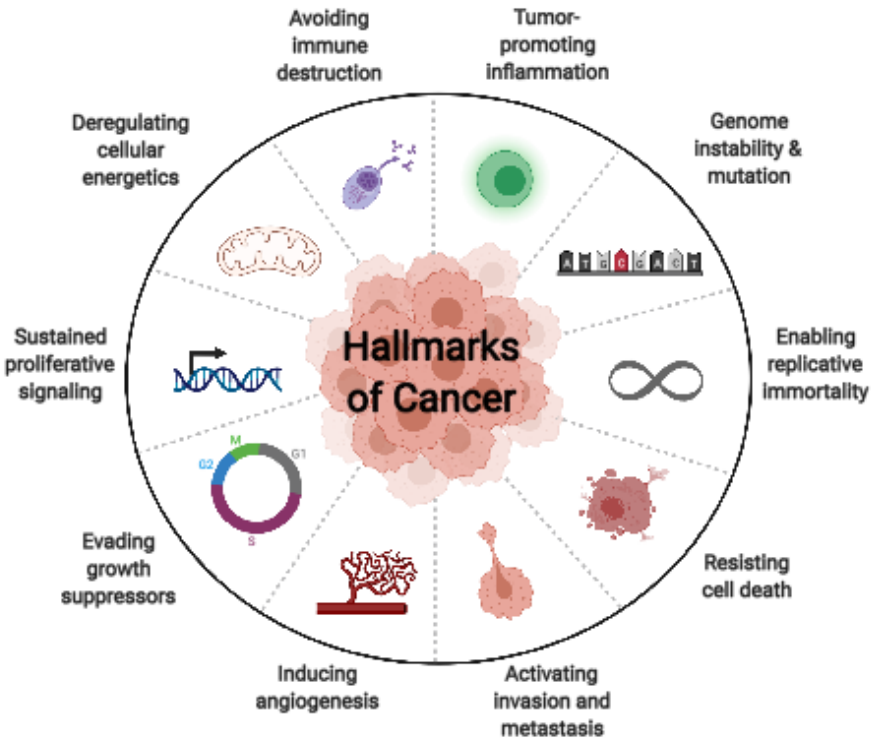

Figure 1.1. Hallmarks of Cancer
Adapted from Hanahan and Weinberg (2011).

Table 1.2. Genetic mutations involved in Colorectal Cancer Development

Tumour Suppressor Genes	Proto-oncogenes	Other genetic or molecular changes
APC	*BRAF*	*BMP3*
ARID1A	*ERBB2*	Chromosomal instability
CTNNB1	*GNAS*	CpG island methylation
DCC	*IGF2*	Microsatellite instability
FAM123B	*KRAS*	Mismatch-repair genes
PTEN	*MYC*	*NDRG4*
RET	*NRAS*	*SEPT9*
SMAD4	*PIK3CA*	*POLD1*
TGFBR2	*RSP02*	*POLE*
TP53	*RSP03*	
	SOX9	
	TCF7L2	

Oncogenes are either mutated to become constantly transcribed or are activate a previously inactive gene (Vicente-Dueñas et al., 2013). Examples of oncogenes in CRC are the *RAS* and *RAF* genes (Section 1.1.3.2.1.4). Most oncogenes are derived from proto-oncogenes that are involved in normal cell growth and proliferation (Torry and Cooper, 1991). Through a GOF mutation, these genes become oncogenic. Unlike TSGs, this activating mutation usually only requires one allele to be mutated for a cancer phenotype to occur, therefore these genes behave in a dominant manner (Knudson, 1971, Knudson, 1996).

1.1.3.2.1 The adenoma-carcinoma sequence

Originally described in the 1970s by studies such as Hill et al. (1978), the adenoma-carcinoma sequence is the step wise action of GOF mutations in oncogenes and LOF mutations in TSG resulting in CRC (Figure 1.3). Initial mutations begin in normal epithelial cells of the colon or rectum, most commonly the crypt (Cernat et al., 2014). Mutations result in the formation of an adenoma, normally through inactivation of *APC* (Fearon and Vogelstein, 1990). Through further mutations of proto-oncogenes and TSGs, these early adenomas progress to intermediate and late adenoma before progression to a carcinoma (Leslie et al., 2002). Throughout this progression, genome stability also decreases, increasing the mutational burden of the tissue (Pino and Chung, 2010).

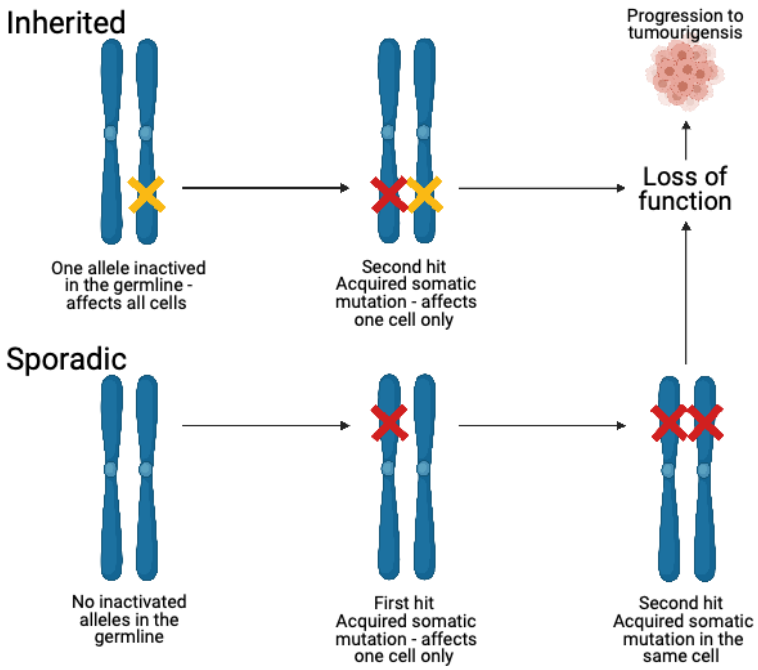

Figure 1.2. Knudson's two-hit hypothesis for loss of function of tumour suppressors in tumorigenesis

In those with inherited disease a germline mutation is found in all cells, therefore only one sporadic mutation is required for loss of function of the gene. In sporadic disease, two somatic mutations must occur (the first and second hit) for the loss of tumour suppressor function and ultimate progression to tumourigenesis. First proposed by Knudson (1971).

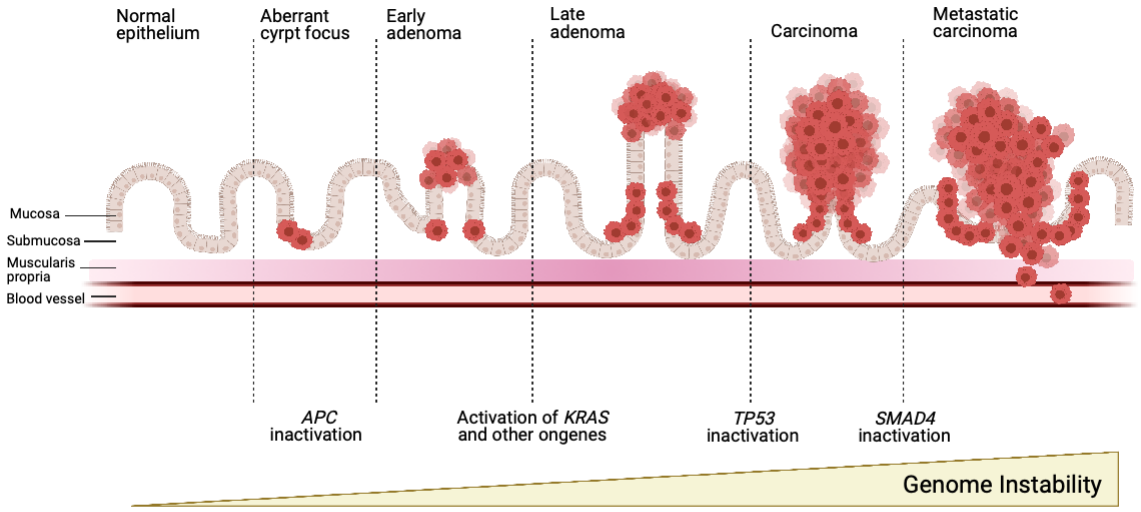

Figure 1.3. Stepwise model of colorectal tumorigenesis
Based on the model proposed by Fearon and Vogelstein (1990). Mutations in genes including *APC*, *RAS* and other oncogenes transform cells from normal epithelium towards malignancy and ultimately metastasis. Throughout this progression of malignancy, genetic instability (including chromosomal instability, microsatellite instability and epigenetic silencing) develops.

1.1.3.3 Genome Instability

A result of mutations in genes associated with DNA repair that occurs throughout the adenoma-carcinoma sequence (Figure 1.3), genomic instability is a characteristic of almost all human cancers – including CRC – and is considered one of the 'hallmarks' of the disease (Figure 1.2) (Negrini et al., 2010, Hanahan and Weinberg, 2011). The most common types of genome instability found in CRC are chromosomal instability (CIN) and microsatellite instability (MSI) which have different mechanisms and implications on the disease.

CIN is found in 70–85% of all diagnosed CRC cases and is described as increased changes in structure and number of chromosomes in cancer cells compared to wild-type tissues due to processes such as abnormal cell cycle progression (Dunican et al., 2002, Walther et al., 2008, Turajlic et al., 2019, Sansregret et al., 2018, Thompson et al., 2010). Increased CIN can result in aneuploidy and increased rates of mutation at areas of CIN (Potapova et al., 2013, Guo et al., 2018). These mutations, if they occur in oncogenes or TSGs, can directly influence colorectal tumorigenesis (Hoevenaar et al., 2020). It has also been shown that the results of this CIN have a direct impact on CRC outcomes (Orsetti et al., 2014).

MSI is found in around 15% of diagnosed CRC cases and is characterised by hypermutation of short repetitive sequences of DNA (1–6 bases repeated up to 50 times) known as microsatellites or short tandem repeats (STRs) (Sinicrope and Sargent, 2012, Nojadeh et al., 2018, Gulcher, 2012, Richard et al., 2008). MSI is due to LOF of mismatch repair mechanisms (Kawakami et al., 2015). It is also characteristic of HNPCC and both inherited and sporadic MSI-positive primary tumours are more likely to be located in the proximal tumour (Lynch and de la Chapelle, 2003, Popat et al., 2005).

A rarer mechanism of genomic instability found in CRCs is as a result of epigenetic silencing events of CpG islands (Dahlin et al., 2010). In some cases, these three types of chromosomal instability have been shown to overlap within the same tumour (Ogino et al., 2009a).

1.1.3.4 The *Adenomatous Polyposis Coli* (*APC*) gene

Outside of CIN, individual genes have been shown to significantly contribute to CRC formation and progression. The most notable of these being the TSG *APC*. Mutations of *APC* result in truncation of the translated protein and are common in cases of both inherited and sporadic CRC (Miyoshi et al., 1992). Germline mutations in *APC* are characteristic of FAP and somatic mutations are found in approximately 80% of sporadic CRC cases (Fearnhead et al., 2001, Galiatsatos and Foulkes, 2006, Cancer Genome Atlas Network, 2012). This indicates that *APC* is a key gatekeeper in CRC progression, with LOF of *APC* occurring early in the adenoma-carcinoma pathway and being a driver mutation of CRC carcinogenesis and metastasis (Powell et al., 1992, Huang et al., 2018). A key molecule found to bind with the *APC* protein is β-catenin, however *APC* is known to influence CRC formation in both a β-catenin dependent and independent manner (Hankey et al., 2018).

The *APC* protein is primarily located in the cytoplasm and therefore influences a number of cellular processes including the wnt-signalling pathway (Neufeld and White, 1997, Schneikert and Behrens, 2007). Wnt-signalling is involved in tissue homeostasis, cell proliferation, polarity and stem cell-fate determination, and dysregulation of the pathway is associated with CRC (Logan and Nusse, 2004, Clevers, 2006, Schatoff et al., 2017). *APC* acts as a scaffold for the complex responsible for the destruction of β-catenin, a key component of the pathway (Graham et al., 2000). Inactivation of *APC* may also reduce cell-to-cell adhesion through its interaction with β-catenin which is involved in the structure of adherens junctions which are linked to the actin cytoskeleton (Bienz and Hamada, 2004, Su et al., 1993). *APC* has also been shown to be involved in cell-cycle progression (Baeg et al., 1995). It has been proposed that this is through the regulation of transcription of S-phase regulators (mediated by β-catenin and Tcf) including c-myc and cyclin D1 (Heinen et al., 2002).

The molecular mechanisms for the *APC* mutations that influence these signalling pathways and DNA repair processes do not entirely follow the classical Knudson ‘two-hit’ hypothesis (Figure 1.1) (Knudson, 1971, Knudson, 1996). In line with Knudson’s hypothesis, LOF of both *APC* alleles occurs at the majority of early CRC tumours,

however *APC* does not entirely follow the model (Miyoshi et al., 1992, Lamlum et al., 1999). Known as the 'just right' hypothesis, a non-random distribution of somatic mutations dependent on the 'first hit' in *APC* may be part of a different process from complete LOF of β-catenin binding motifs in *APC* (Lamlum et al., 1999). There is evidence of gene mutations in specific areas to maintain optimum levels of β-catenin accumulation, with too low only allowing for access to the most responsive genes and too high expression resulting in induction of cell death and the destruction of the cancer cells (Albuquerque et al., 2002, Romagnolo et al., 1999, Kim et al., 2000). This 'just-right' nature of *APC* expression is only one part of the balance of expression of a number of genes required for CRC progression.

1.1.3.5 Genes associated with the Epidermal Growth Factor Receptor (EGFR) pathway

EGFR, a transmembrane protein that is a member of the ErbB family of receptors, is the key intracellular component for which the *EGFR* signalling pathway is named after (Zhang et al., 2007). EGFR is a receptor for members of the epidermal growth factor (EGF) family of ligands and the resulting signalling has been shown to be one of the most important pathways for regulation of growth, proliferation, differentiation and survival of cells (Oda et al., 2005, Cohen et al., 1980). The dysregulation of the EGFR pathway is well established in cases of epithelial cancer and genes or proteins found within it have become well-established biomarkers for treatment response (Sigismund et al., 2018). Between 60 and 80% of diagnosed CRC cases have an element of EGFR up-regulation (Goldstein and Armin, 2001).

Activation of the EGFR signalling pathway requires ligand binding to the receptor. Whilst eight ligands are known to bind to *EGFR*, the main focus of research and understanding of the pathway has been focused on EGF and transforming growth factor α (TGF- α) (Henriksen et al., 2013). *EGFR* can also be activated in a ligand independent manner with a variety of signalling outcomes (Guo et al., 2015).

In both ligand and ligand-independent EGFR activation, different pathways and molecules are activated, with over 219 molecules being either directly or indirectly

influenced by EGFR (Oda et al., 2005). However, the most explored in relation to CRC are the RAS-RAF-MEK-ERK pathway (also referred to as the mitogen-activated protein kinase (MAPK) pathway) and PI3K/AKT pathway (Figure 1.4). Increased activation of these pathways has been shown to influence the growth of tumours (Fang and Richardson, 2005, Danielsen et al., 2015, Yarden, 2001).

Mutations in the genes that encode the proteins involved in these signalling pathways have been shown to negatively impact survival and other cancer outcomes, both when patients are treated with a therapy designed to directly target this pathway and other therapeutics which are not molecularly targeted, such as cytotoxic therapies (Gong et al., 2016, Phipps et al., 2013, Mei et al., 2016). Most notable is the impact of somatic mutations including *RAS* mutations (*KRAS* and *NRAS*) – which are mutated in around 45% of all CRC tumours – on specific anti-EGFR therapies (Section 1.1.4.2) (Smith et al., 2013b, Douillard et al., 2013, Lièvre et al., 2006).

1.1.4 CRC treatment

Treatment for CRC has undergone major developments since the introduction of 5-Flurouracil (5-FU) in the early 1960s (Table 1.3); this has resulted in increased survival in CRC patients and changes in clinical practice (Van Cutsem et al., 2016).

Current clinical practice employs at least one, or a combination of, surgery, radiotherapy, chemotherapy and molecularly targeted therapies (Kuipers et al., 2015). The most common first-line treatment for CRC is resection of the tumour(s) through surgery which can occur at all stages of diagnoses (Ghiasloo et al., 2020, Rentsch et al., 2016). Radiation can be used in conjunction with other treatments including prior to surgery in order to decrease tumour size (Higgins et al., 1975). However, the two treatments that have arguably increased survival the most and are therefore the main focus of clinical trials for CRC (Table 1.4) are cytotoxic and molecularly targeted therapies.

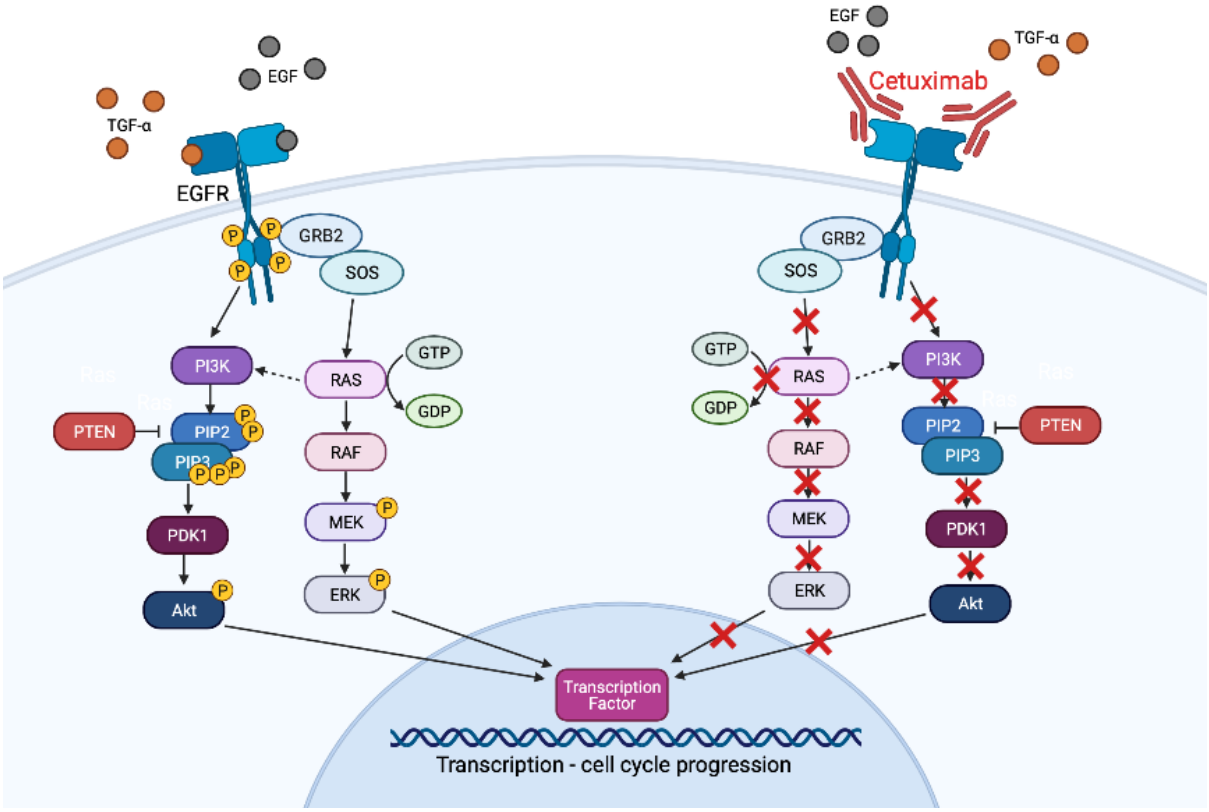

Figure 1.4. EGFR signalling pathway

Ligands (including EGF and TGF-α) bind to the EGFR receptor, which causes activation via phosphorylation and triggering of cellular signalling pathways. These pathways contribute to transcription of genes which promote cell cycle progression. When an anti-EGFR monoclonal antibody (mAb; such as cetuximab) is present, the ligand is unable to bind and activation of this signalling cascade does not occur (indicated by red crosses). Should an activation mutation occur in a component of the pathway, such as *KRAS* or *BRAF*, the pathway activates from the mutated point onwards, rendering the mAb ineffective.

Table 1.3. Treatments for colorectal cancer

Therapy	Description	Use	Year	Approval or Advice
5-Fluorouracil	Thymidylate synthase (TS) inhibitor	Chemotherapy treatment for CRC	1962	FDA
Adjuvant therapy	Therapy given in addition to an initial therapy to increase efficacy	Chemotherapy as a secondary treatment following surgical resection	1990	
Irinotecan	Cytotoxic alkaloid which inhibits topoisomerase I	In conjunction with FOLFIRI as a first-line treatment	1996	FDA (accelerated)
		Second-line monotherapy	1998	FDA (full) EMA
Oxaliplatin	Platinum-based compound with a non-targeted cytotoxic action	In conjunction with 5-FU (FOLFOX) as a second-line treatment	1996	EMA (initially France)
			2002	FDA
Capecitabine	5-FU precursor	Oral therapy used in conjunction with oxaliplatin (XELOX) or irinotecan (XELIRI)	2001	FDA EMA
Bevacizumab	mAB – vascular endothelial growth factor A (VEGF-A) inhibitor	First-line treatment for standard chemotherapy treatment	2004	FDA
		Second-line treatment in conjunction with 5-FU chemotherapy (eg XELOX or FOLFOX)	2005 2006	EMA
Cetuximab	mAB – epidermal growth factor receptor (EGFR) inhibitor	Monotherapy Combination therapy with irinotecan	2004	FDA & EMA
		Combination therapy with FOLFOX in previously untreated patients	2009	FDA
		Only patients with *KRAS* wild-type tumours recommended for treatment	2010	EMA
		Only patients with *RAS* wild-type tumours recommended for treatment		
Pantimumab	mAB -– epidermal growth factor receptor (EGFR) inhibitor	Monotherapy	2006	FDA
		First-line treatment in combination with FOLFOX Second-line treatment in combination with FOLFIRI	2007	EMA
		Only patients with *KRAS* wild-type tumours recommended for treatment	2009 2010	FDA EMA

		Only patients with *RAS* wild-type tumours recommended for treatment		
Regorafenib	Multi-kinase inhibitor	Oral therapy after lack of response to other approved therapies	2012 2013	FDA EMA
Aflibercept	Vascular endothelial growth factor A (VEGF-A) and placental growth factor (PIGF) antagonist	Combination therapy with FOLFIRI	2012 2013	FDA EMA
Trifluridine/tipiracil (TAS-102)	Cytotoxic pyrimidine and thymidine phosphorylase (TP) inhibitor	Oral therapy after lack of response to other approved therapies	2015 2016	FDA EMA
Encorafenib	Small molecule ATP-competitive BRAF kinase inhibitor	Oral therapy for patients with *BRAF* V600E mutant colorectal tumours	2018 2020	EMA FDA

EMA=European medicines agency, FDA=food and drug administration, FOLFIRI=folinic acid, fluorouracil and irinotecan, FOLFOX=folinic acid, fluorouracil and oxaliplatin, mAb=monoclonal antibody, XELIRI=Irinotecan and capecitabine, XELOX=oxaliplatin and capecitabine.

Table 1.4. Clinical trials examining commonly used CRC treatments

Trial name	Published study	Treatment	Patients recruited	Primary outcomes reported
COX suppressors				
ADD-ASPIRIN	Coyle et al. (2016)	Aspirin	~2,600	In progress
VICTOR	Pendlebury et al. (2003) Midgley et al. (2010)	Rofecoxib	2,434	NSD in OS (HR=0.97, *P*=0.75)
Cytotoxic chemotherapy				
CAIRO	Koopman et al. (2007)	Capecitabine, Irinotecan, Oxaliplatin	820	NSD in OS between sequential and combination groups (HR=0.92, *P*=0.33)
CORGI-L	Gunnlaugsson et al. (2009)	XELOX	47	78% alive – no control arm
FOCUS	Seymour et al. (2007)	Fluorouracil, Irinotecan, Oxaliplatin	2,135	NSD between treatments (HR=1.06, *P* not reported)
FOCUS2	Seymour et al. (2011)	Capecitabine, Fluorouracil, Oxaliplatin	459	Quality of life analyses – Overall Treatment Utility
MOSAIC	André et al. (2004) André et al. (2009)	Fluorouracil, Levamisole, Oxaliplatin	2,246	↑ DFS in combination therapy (HR=0.80, *P*<0.01)
QUASAR	Gray et al. (2007)	Fluorouracil and folonic acid	3,239	↓ RR of death (HR=0. 78, *P*<0.01)
SCOT	Iveson et al. (2018)	FOLFOX, XELOX	6,088	NSD in PFS between delivery times (HR=1.01, Non-inferiority *P*=0.01)
TRANSSCOT	Engelmann et al. (2016)	FOLFOX, XELOX	6,144	Translational study
Kinase Inhibitor				
CORRECT	Grothey et al. (2013)	Regorafenib	760	↑ OS (HR=0.77, *P*<0.01)
FOCUS4-D	Adams et al. (2018)	Tyrosine kinase inhibitor AZD8931	32	NSD in OS between groups (HR=1.10, *P*=0.95)
Monoclonal Antibody				
BEACON CRC	Van Cutsem et al. (2018) Kopetz et al. (2019)	Binimetinib, Cetuximab,	665	↑ OS in triple therapy (HR=0.52, *P*<0.01)

		Encorabenib, FOLFIRI		
CAIRO2	Tol et al. (2008) Tol et al. (2009)	Bevacizumab, Capecitabine, Cetuximab, Oxaliplatin	529	↓ PFS on cetuximab (HR=1.22, *P*=0.01)
CAIRO3	Simkens et al. (2015)	Bevacizumab, Capecitabine, Oxaliplatin	558	↑ PFS (HR=0.67, *P*<0.01)
CAIRO4	t Lam-Boer et al. (2014)	Bevacizumab, Fluoropyrimidine-based chemotherapy	~350	In progress
CAIRO5	Huiskens et al. (2015) Huiskens et al. (2019)	Bevacizumab, FOLFIRI, FOLFOX, Panitumumab	~650	In progress
COIN	Adams et al. (2011) Maughan et al. (2011)	Cetuximab, XELOX, FOLFOX	2,445	NSD in OS between continuous and intermittent chemotherapy (HR=1.09, *P* not reported) NSD in OS with cetuximab (HR=1.01, *P*=0.87)
COIN-B	Wasan et al. (2014)	Cetuximab, FOLFOX	226	NSD in OS between intermittent or maintained cetuximab (HR and *P* not reported)
CRYSTAL	Van Cutsem et al. (2009) Van Cutsem et al. (2011)	Cetuximab, FOLFIRI	1,198	↑ PFS in cetuximab group (HR=0.85, *P*=0.05) NSD in OS between groups (HR=0.93, *P*=0.31)
EXCITE	Gollins et al. (2011) Gollins et al. (2017)	Capecitabine, Cetuximab, Irinotecan	82	88.2% OS.
FOCUS3	Maughan et al. (2014)	Bevacizumab, Cetuximab, Fluorouracil, Irinotecan, Oxaliplatin	240	In progress
FOCUS4	Kaplan (2015)	Aspirin, Capecitabine, Panitumumab, other novel agents	~4,700	In progress

FOxTROT	Foxtrot Collaborative Group (2012) Morton (2019)	FOLFOX, Panitumumab, XELOX	1,052	↓ 2 year failure rate (HR=0.77, *P*=0.11)
ICE CREAM	Segelov et al. (2016)	Cetuximab, Irinotecan	100	In progress
OPUS	Bokemeyer et al. (2011)	Cetuximab, FOLFOX	315	↑ PFS on cetuximab in *KRAS* wild-type patients (HR=0.57, *P*<0.01)
PICCOLO	Seymour et al. (2013)	Ciclosporin, Irinotecan, Panitumumab	1,198	↑ PFS on panitumumab in *KRAS* wild-type patients (HR=0.78, *P*=0.02) NSD in OS (HR=1.01, *P*=0.91)
PRIME	Douillard et al. (2010) Douillard et al. (2014)	FOLFOX, Panitumumab	1,183	↑ PFS on panitumumab in *KRAS* wild-type patients (HR=0.80, *P*=0.01) ↑ OS on panitumumab in *KRAS* wild-type patients (HR=0.83, *P*=0.03)
QUASAR 2	Rosmarin et al. (2014) (Kerr et al., 2016)	Bevacizumab, Capecitabine	1,952	NDS in DFS (HR=1·06, *P*=0·54)
SOFT	Yamada et al. (2013) Baba et al. (2017)	Bevacizumab, FOLFOX, SOX	512	↑ PFS on SOX (HR=1.05, non-inferiority *P*=0.01)
Nucleoside analogue				
RECOURSE	Mayer et al. (2015) Van Cutsem et al. (2018)	Lonsurf	800	↑ OS (HR=0.68, *P*<0.01)

↑=Significantly increased, ↓=Significantly decreased, DFS=Disease-free survival, FOLFIRI=folinic acid, fluorouracil and irinotecan, FOLFOX=folinic acid, fluorouracil and oxaliplatin, HR=Hazard Ratio, NSD=No significant difference OS=Overall survival, RR=Risk ratio, PFS=Progression-free survival, SOX=S-1 and oxaliplatin, XELOX=oxaliplatin and capecitabine

1.1.4.1 Cytotoxic therapies

Cytotoxic therapies, chemotherapies which encourage cell death, are derived from gases used in the first and second world war (Kon and Ross, 1948). Whilst effective, a mechanism of action which targets essential biological processes results in diverse toxic side effects across multiple organ systems (Kummar et al., 2006). Despite this, cytotoxic chemotherapy is still a commonly used CRC treatment (Kuipers et al., 2015).

1.1.4.1.1 Fluoropyrimidines

In mCRC and aCRC, the common molecule for first-line palliative therapy is fluoropyrimidine (FP). This can either be in the form of intravenous 5-FU or the oral 5-FU precursor capecitabine which is metabolised in a manner that mimics continuous infusion of 5-FU (Vodenkova et al., 2020, Chintala et al., 2011, Van Cutsem et al., 2004, Johnston and Kaye, 2001). FPs can either be given as monotherapy or in combination with other cytotoxic therapies (Table 1.3). Notable combination therapies include a combination of intravenous 5-FU, folinic acid (leucovorin) and irinotecan (FOLFIRI); intravenous 5-FU, leucovorin and oxaliplatin (FOLFOX or oxaliplatin modified de Gramont; OxMdG) and a combination of orally administered capecitabine and intravenous oxaliplatin (XELOX or CAPOX) (Cassidy et al., 2004, Van Cutsem et al., 2009, Pasetto et al., 2005).

5-FU is an ‘antimetabolite drug’ that inhibits cancer growth through inhibition of essential biological processes (including synthesis of essential macromolecules). In the case of 5-FU, the mechanisms of action are thymidylate synthase (TS) inhibition and RNA misincorporation (Longley et al., 2003).

Thymidylate synthase (TS) catalyses the conversion of deoxyurudine monophosphate (dUMP) into deoxythmidine monophosphate (dTMP) - a nucleoside (thymine) phosphate (Rose et al., 2002, Wilson et al., 2014). dTMP is then phosphorylated into deoxythymidine triphosphate (dTTP), an essential precursor for the DNA synthesis (Chu et al., 2003). This pathway is the only *de novo* source of dTTP, therefore inhibition of TS (by anticancer drugs such as 5-FU) causes deoxynucleotide imbalance

which are believed to disrupt DNA synthesis causing lethal DNA damage (Brandt and Chu, 1997, Houghton et al., 1995, Yoshioka et al., 1987).

5-Fluorouridine triphosphate (FTP), a 5-FU metabolite, is frequently misincorporated into RNA resulting in disruption of normal RNA processing and function (Longley et al., 2003). This misincorporation can result in RNA toxicity through inhibition of pre-rRNA processing, disruption of post-translation modifications of tRNAs and inhibition of splicing of pre-mRNA (Kanamaru et al., 1986, Santi and Hardy, 1987, Doong and Dolnick, 1988).

1.1.4.1.2 Oxaliplatin

Two of the chemotherapeutics used in the COIN and COIN-B clinical trials contain oxaliplatin, which can also be used as a single agent (Rothenberg, 2000, Wasan et al., 2014, Adams et al., 2011, Maughan et al., 2011). Sold under the brand name Eloxatin, it is a novel platinum-based compound with a bidentate ligand trans-1,2-diaminocyclohexane and a bidentate oxalate group (Apps et al., 2015). This bidentate oxalate group, sometimes referred to as the “leaving group”, is responsible for some oxaliplatin’s unique pharmacokinetic profile (Raymond et al., 1998). This includes the non-enzymatic transformation into other reactive compounds through oxalate group displacement that occurs in plasma (Alcindor and Beauger, 2011).

One method of oxaliplatin’s action is cytotoxicity through DNA damage and disruption of DNA synthesis. Oxaliplatin induces DNA lesions through the formation of intra- and inter-strand DNA crosslinks and the formation DNA-protein crosslinks (Faivre et al., 2003, Zwelling et al., 1979). There is evidence that the formation of these DNA crosslinks results in apoptosis (Faivre et al., 2003).

Oxaliplatin also inhitibits the synthesis of mRNA (transcription). The three main mechanism of this inhibition are direct binding of transcription factors and platinum-DNA molecules and inhibition of RNA polymerases due to the inability of platinum bound DNA to enter enzyme active sites (Todd and Lippard, 2009).

Interestingly, oxaliplatin has also been shown to have an antimetabolite-like effect on TS causing arrest of mitosis (Fischel et al., 2002). However, as oxaliplatin is combined with 5-FU, it is unclear if this mechanism of action is significant *in vivo* alone (Pasetto et al., 2005).

Oxaliplatin-based chemotherapeutics, like 5-FU, target CRC in a cytotoxic manner which results in serious side effects particularly in the gastrointestinal, hematopoietic and peripheral nervous systems (Alcindor and Beauger, 2011, Graham et al., 2004).

1.1.4.2 Molecularly targeted therapies

In conjunction with, or independent of, cytotoxic chemotherapies, molecularly targeted therapies have been shown to improve mCRC outcome (Van Cutsem et al., 2014). These therapies are commonly the subject of clinical trials into new treatments (Table 1.4) some which are routinely adopted in clinical practice (Table 1.3).

Molecularly targeted therapies interact cancer associated proteins or those that are involved in pathways related to cell proliferation, progression and tumorigenesis (Padma, 2015, Pérez-Herrero and Fernández-Medarde, 2015). These therapies can be separated into three main categories - immunotoxins, small molecule inhibitors and mABs (Baudino, 2015). In relation to the work in this thesis, the most notable of these are mAbs.

mAbs target cell receptors dysregulated in CRC. These include the vascular endothelial growth factor (VEGF) signalling pathway (bevacizumab) and the EGFR pathway (cetuximab) (Ranieri et al., 2006, Galizia et al., 2007). Cetuximab binds to the EGFR receptor (Figure 1.4) preventing ligands such a EGF and TGF-α from binding, inhibiting downstream signalling (Section 1.1.3.5) and reducing proliferation that is characteristic of cancer progression (Smith et al., 2013b). Side effects from these treatments are more specific than those found in cytotoxic therapies (which can inhibit highly proliferative healthy cells alongside cancer cells) and include hypomagnesaemia and allergic responses (Wolpin and Mayer, 2008, Pérez-Herrero and Fernández-Medarde, 2015).

Whilst there is evidence that cetuximab can increase OS, this is not the case for all mCRC patients. Patients with somatic mutations in key genes in the EGFR pathway, such as *KRAS* and *NRAS*, have been shown to have no benefit from the treatment (De Roock et al., 2010a, De Roock et al., 2010b). The GOF mutations of *RAS* cause the downstream activation of a signalling cascade that will cause uncontrolled proliferation, overcoming the upstream activity of cetuximab (Karapetis et al., 2008, De Roock et al., 2010a). This means that only patients with *RAS* wild-type tumours are given cetuximab in the clinic (NHS National Institute for Health and Clinical Excellence (NICE), 2017, Van Cutsem et al., 2016).

1.1.4.3 Clinical trials for CRC treatments

The efficacy of new treatments are tested to determine the most effective treatment of CRC by clinical trials (Table 1.4). These trials use different primary endpoints including survival measures (such as OS and progression-free survival, PFS) and tumour response, using standardised measures such as the RECIST guidelines (Korn et al., 2011, Therasse et al., 2000). Some trials also collected blood samples or formalin-fixed, paraffin-embedded (FFPE) tumour samples from resection to perform translational studies to explore the underlying genetic patterns that underpin these differing outcomes (Maughan et al., 2011, Smith et al., 2013a).

My investigations used tumour and blood samples from advanced CRC patients recruited for the MRC clinical trials – COIN (ISRCTN27286448) and COIN-B (ISRCTN3837568, Sections 1.1.4.4 and 1.1.4.5).

1.1.4.4 COIN

COIN (COntinuous versus INtermittent) was a Cancer Research UK and MRC funded phase III clinical trial. Two thousand, four hundred and forty-five patients with aCRC recruited by consultant oncologists from centres in the UK and Ireland between 2005 and 2008. Inclusion criteria included informed written consent, ≥18 years old, histologically confirmed adenocarcinoma of the colon or rectum, inoperable locoregional or metastatic disease as described by RECIST version 1.0, good end-organ function and WHO performance status 0–2 (Therasse et al., 2000). Patients

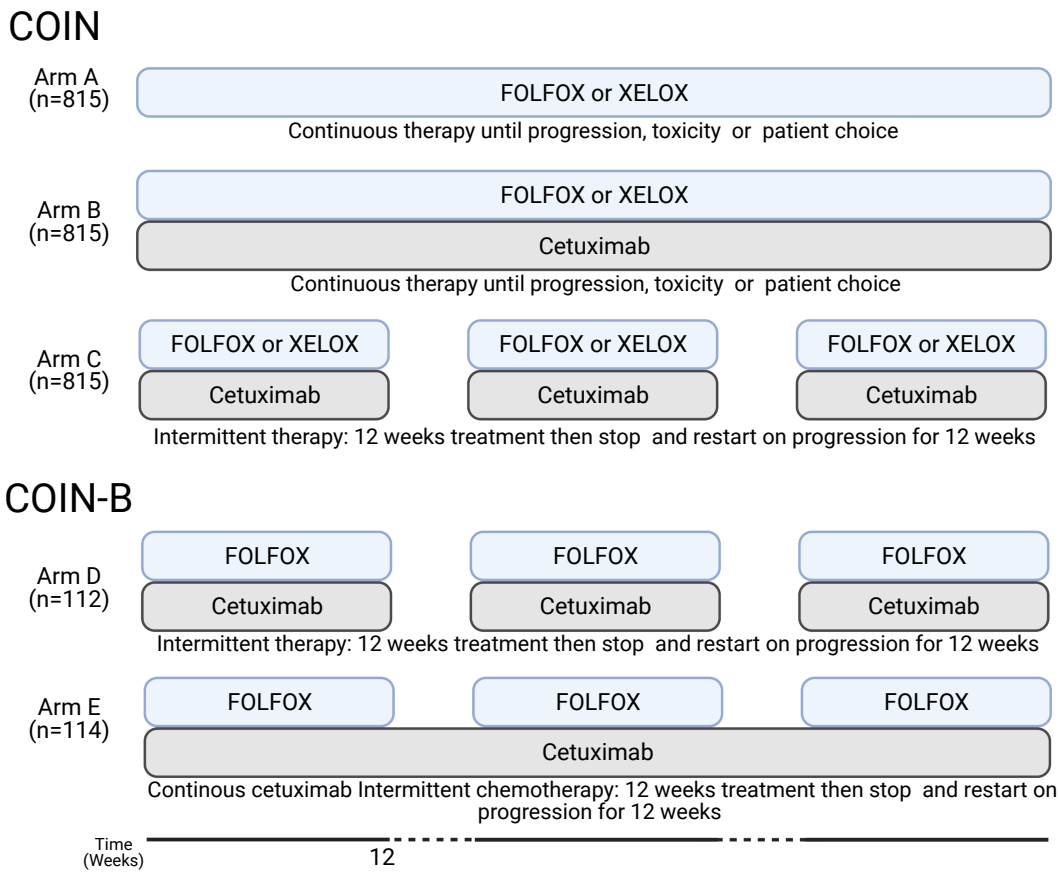

Figure 1.5. COIN and COIN-B trial design
Adapted from Adams et al. (2011) and Wasan et al. (2014). On the intermittent arms, treatment was given for 12 weeks, stopped and resumed on progression for 12 weeks.

were excluded based on identified brain metastases, previous or present malignant disease, previous exposure to oxaliplatin or uncontrolled medical comorbidity that may interfere with treatment or assessment of predefined endpoints (Maughan et al., 2011).

Patients were randomised into three arms (1:1:1) and received either continuous oxaliplatin and fluoropyrimidine combination (Arm A, n=815, Figure 1.5), cetuximab and continuous oxaliplatin and fluoropyrimidine combination (Arm B, n=815) or intermittent oxaliplatin and fluoropyrimidine combination (Arm C, n=815). Patients were given 5-FU and oxaliplatin (FOLFOX) or oral capecitabine and oxaliplatin (XELOX) prior to randomisation. One third were given FOLFOX with the remaining patients receiving XELOX (Smith et al., 2013c). Cessation of arms A and B occurred on patient choice, disease progression or toxicity (Adams et al., 2011).

COIN was designed to determine whether there was significant difference in outcome between patients who received continuous or intermittent chemotherapy (Arm A vs C) and investigate the effect of the addition of the mAb cetuximab to continuous chemotherapy on survival (OS and PFS; Arm A vs B).

Intermittent chemotherapy was shown to be non-detrimental when compared to continuous chemotherapy alone. There was no significant benefit for continuous therapy (however, there is a non-significant (P<0.05) trend towards a benefit of continuous treatment) either in relation to OS (ITT population median OS 15.8 months continuous chemotherapy, 14.4 months intermittent chemotherapy; HR=1.08, 95% CI 0.97–1.21, P not reported; Figure 1.6A) or strategy-failure-free (SFFS) survival (ITT population median SFFS 8.4 months continuous chemotherapy, 7.4 months intermittent therapy; HR=1.05, 95% CI 0.95–1.17, P not reported). Exploratory analyses of clinicopathological factors was performed based on OS data. Patients with *KRAS* wild-type tumours favoured continuous therapy over intermittent therapy (HR=1.23, 95% CI=0.99–1.54, P not reported) although the interactive effect of tumoural *KRAS* mutation and treatment delivery was not statistically significant ($P_{Interaction}$=0.07) (Adams et al., 2011).

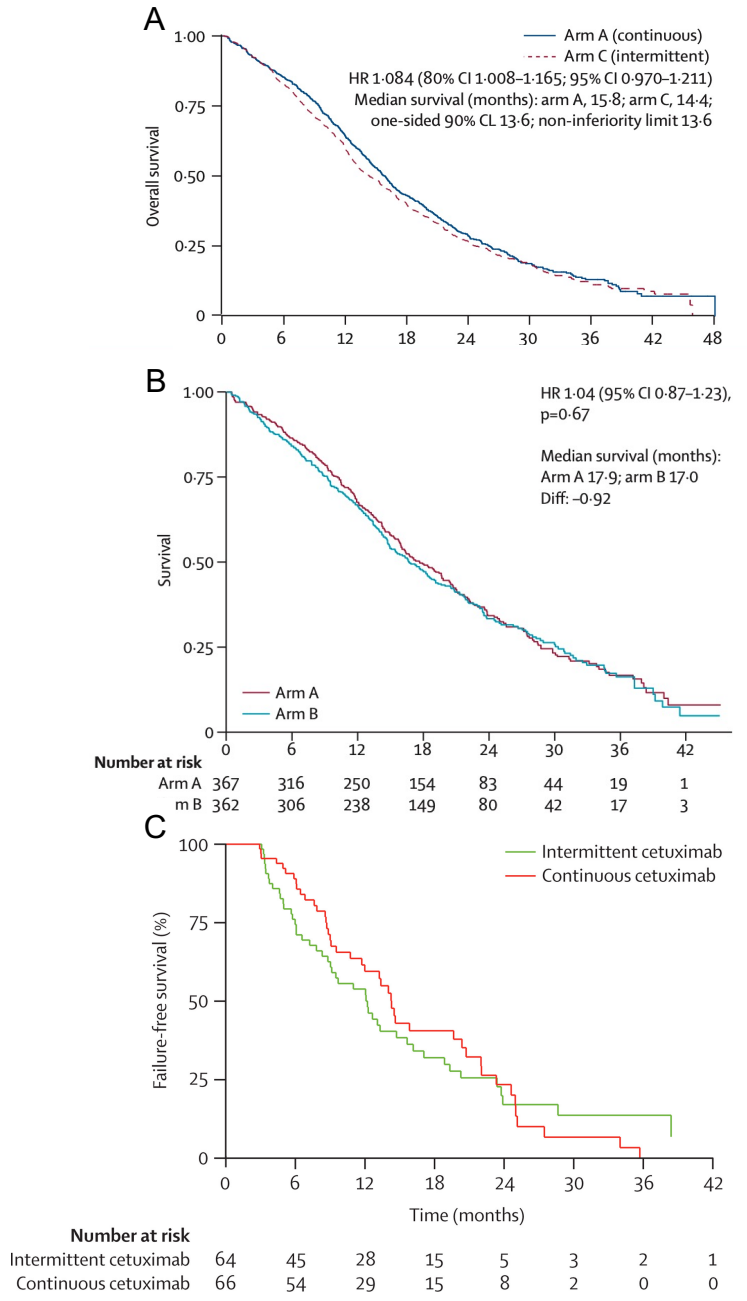

Figure 1.6. Survival analyses from the original COIN and COIN-B trials
Kaplan-Meier curves for overall survival between: (A) continuous and intermittent chemotherapy in the intention to treat population in the COIN trial (B) continuous chemotherapy and continuous chemotherapy plus cetuximab in *KRAS* wild-type patients the COIN trial and (C) continuous and intermittent cetuximab in *KRAS* wild-type patients in the COIN-B trial. Original figures from Adams et al. (2011), Maughan et al. (2011) and Wasan et al. (2014). Reproduced under Creative Commons CC-BY license and for Figure 1.6C permission from Elsevier Global Rights Department.

There was no evidence of cetuximab benefit in patients recruited to COIN. Addition of cetuximab resulted in no significant different in OS (median OS 15.8 months no cetuximab, 15.3 months cetuximab; HR=1.01, 95% CI=0.90–1.13, *P*=0.87) and PFS (median PFS 8.1 months no cetuximab, 7.9 months cetuximab; HR=0.98, 95% CI=0.89–1.09, *P*=0.98) when comparing all recruited patients. There was also no evidence for cetuximab benefit in patients with *KRAS* wild-type CRCs (median OS 17.9 months no cetuximab, 17.0 months cetuximab; HR=1.04, 95% CI=0.87–1.23, *P*=0.67; Figure 1.6B) or PFS (median 8.6 months for both cetuximab and no cetuximab; HR=0.96, 95% CI=0.82–1.12, *P*=0.60).

1.1.4.5 COIN-B

COIN-B, performed as an adjunct to COIN, was an MRC funded Phase II trial. Two hundred and twenty-six aCRC patients were recruited from 30 hospitals in the UK and one in Cyprus between 2008 and 2010. The trial was designed prior the emergence of data showing that tumoural *KRAS* mutations were predictive of resistance to anti-EGFR targeted therapies (Karapetis et al., 2008, Lièvre et al., 2006, De Roock et al., 2008, Amado et al., 2008). Based on this information, trial recruitment was suspended in May 2008. Prior to recommencement in January 2009, tumoural *KRAS* mutation status of already recruited patients was assessed and the protocol for future recruitment was amended to include prospective screening of tumoural *KRAS* mutation status (Wasan et al., 2014).

Other eligibility criteria included informed written consent, ≥18 years old, inoperable locoregional or metastatic colorectal adenocarcinoma as defined by RECIST version 1.1, good organ function and WHO performance status 0–2 (Eisenhauer et al., 2009). Exclusion criteria included uncontrolled comorbidity that may interfere with treatment or assessment of clinical endpoints, any previous cancer or known metastases to the brain.

COIN-B patients were randomised 1:1 to receive intermittent FOLFOX and cetuximab (Arm D, n=112, Figure 1.5) or intermittent FOLFOX and continuous cetuximab (Arm E, n=114). Each group received FOLFOX and weekly cetuximab for 12 weeks followed

by interruption of FOLFOX (and cetuximab for Arm D) until RECIST progression whereupon the same 12-week treatment recommenced.

COIN-B was designed to investigate the potential inferiority of intermittent cetuximab in comparison to continuous cetuximab patients in relation to survival outcomes (OS and PFS; Arm D vs E). Results for 169 *KRAS* wild-type patients in the primary analysis showed benefit from continuous treatment (median PFS intermittent cetuximab 3.1 months, 95% CI=2.8–4.7; continuous cetuximab 5.8 months, 95% CI=4.9–8.6) and failure-free survival (FFS) (FFS intermittent cetuximab 16.8 months, 95% CI=14.5–22.6; continuous cetuximab 22.2 months, 95% CI=18.4–28.9; Figure 1.6C, *P*-values not reported).

1.2 The immune system

The immune system is a complex network of biological molecules and processes that protect an organism from external organisms and objects. The role of the immune system can be split into three broad purposes – fight pathogens entering the body; recognise and neutralise harmful foreign environmental substances and detect and destroy disease-causing changes to cells within the body – some of which may result in cancer phenotypes (Roberts, 2015).

The immune system can be divided into two major components – innate and adaptive. Innate immunity consists of a widely non-specific first-line of defence whereas the adaptive system is highly specialised processes geared towards the elimination of specific threats (Riera Romo et al., 2016, Bonilla and Oettgen, 2010). Whilst these systems are conceptually distinguishable (Table 1.5), they heavily influence each other with overlapping cell types (Figure 1.7) and functionally (Basile et al., 2017, Marshall et al., 2018). A key example of this is dendritic cells (DCs), which induce primary tumour response through the capture and processing of antigens which are then presented to adaptive immune cells, causing polarisation of these to effector cells (Shortman and Liu, 2002).

1.2.1 Role of immunity in CRC

Immune activity has been established to be a ‘hallmark of cancer’ (Figure 1.1), with components of the immune system being shown to contribute to the development and prognosis of CRC (Hanahan and Weinberg, 2011, Markman and Shiao, 2015). There is evidence that changes – including expression levels and mutations – in genes associated with the immune system may impact the risk of developing CRC (Catalano et al., 2019). As a result of a GWAS to discover risk alleles for CRC, gene-set enrichment analysis has shown an enrichment of TGF-B signalling and other immune pathways in patients greater at risk of disease development (Law et al., 2019).

Table 1.5. Characteristics of the innate and adaptive immune system

	Innate	Adaptive
Specificity	Non-specific pattern recognition	Specific recognition of antigens
Activation time	Immediate	Delayed
Recognition	Pattern-associated molecular patterns (PAMPs), Damage-associated molecular patterns (DAMPs)	Specific antigens
Receptors	Formyl peptide receptors, Toll-like receptors, Nod-like receptors, C-type lectin receptors	B-cell receptor, T-cell receptor
Soluble effectors	Complement system, acute-phase reactants	Antibodies
Cellular effectors	Neutrophils, Macrophages and myeloid cells	T-and B-lymphocytes
Diversity	Non-anticipatory that is the same for each exposure	Anticipatory immunity enhanced by antigen exposure
Memory	Low	High

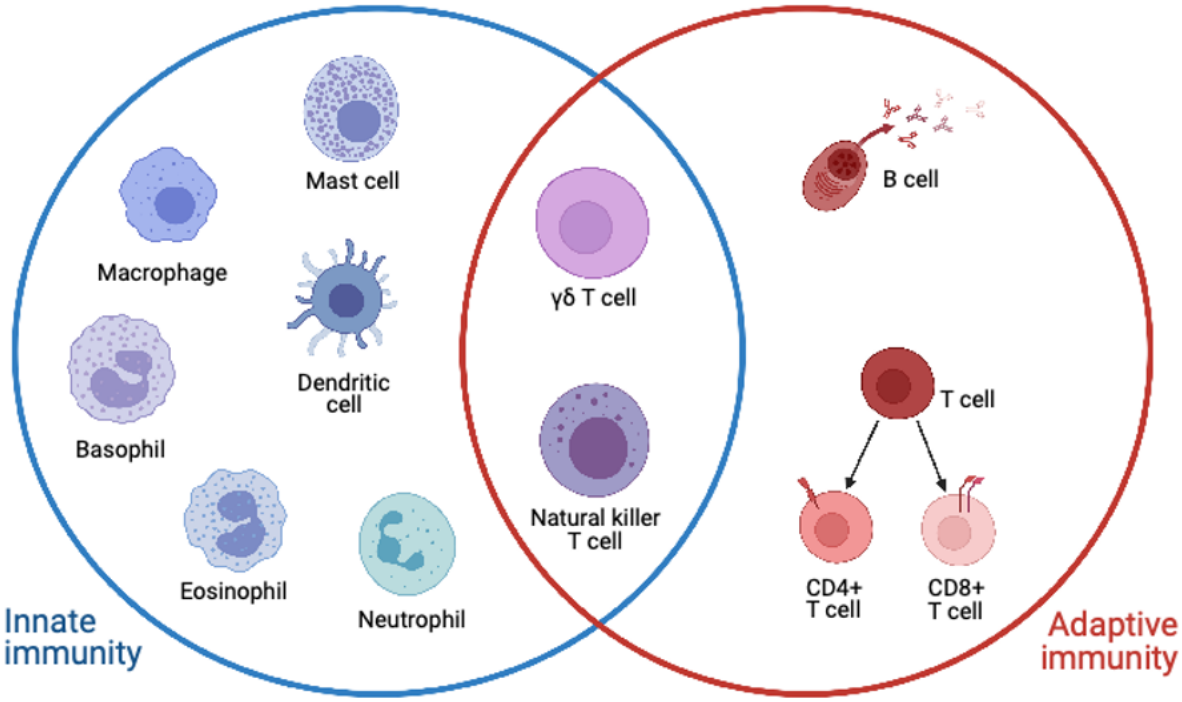

Figure 1.7. Cells of the innate and adaptive immune system
Innate immune cells are coloured in blue, adaptive immune cells are red and those shared by both immune cells purple.

The antitumour immune response impacts on the clinical outcome of all stages of CRC (Markman and Shiao, 2015). It has been established that cytotoxic chemotherapies such as oxaliplatin, utilised in the treatment of CRC, can trigger immunogenic cell death in cancer cells (Obeid et al., 2007).

1.2.1.1 Adaptive immunity and colorectal cancer

Research into the antitumour response has mainly focused on the components of the adaptive immune system (Munhoz and Postow, 2016). The most notable component of the adaptive immune system which has been shown to impact the antitumour response is the activity of T-cells (Kennedy and Celis, 2008, Tay et al., 2021, Tsukumo and Yasutomo, 2018). These specialised cells play a key role in CRC and can have an impact on tumourigenesis (Nosho et al., 2010, Mueller and Fusenig, 2004). Increased T-cell activity is associated with a better prognosis in CRC (Galon et al., 2006, Nosho et al., 2010). Therefore, the presence of different types of T-cells influence colorectal tumour formation, progression and prognosis.

Cytotoxic T-lymphocytes ($CD8^+$ T-cells) are a leading component of antitumour immunity. They recognise tumour antigens on presented human leukocyte antigen (HLA) class I proteins (Paschen et al., 2004). The resulting protein complex results in expansion of CD8+ T-cells causing their activation (Brazin et al., 2015). These cells then mediate the specific destruction of tumour cells through regulation of destructive enzymes and activation of apoptosis (Siska and Rathmell, 2015). Levels of CD8+ T-cell activity is therefore a prognostic marker for survival, with cells at the centre of the tumour having the most significant impact on OS (Idos et al., 2020, Waldner et al., 2006).

CD4+ T-cells (or helper T-cells) interact with antigens by HLA Class II proteins presented on DCs, have also been shown to influence tumour response (Hammer and Ma, 2013, Campi et al., 2003). CD4+ T-cells are highly versatile and have varied roles in tumour development, antitumour response and CRC prognosis. They have been shown to differentiate into many subtypes with specific functions that ‘help’ to coordinate immune response via recruitment of effector immune cells (Crotty, 2014,

Tay et al., 2021). A subset of CD4+ T-cells – Th17 cells, which are known to produce the cytokine IL-17 – have been shown to have a prognostic effect on CRC (Niccolai et al., 2017, Tosolini et al., 2011, Nosho et al., 2010, Pages et al., 2005, Korn et al., 2009). This cytokine may have a negative impact on OS and DFS in CRC patients, key markers used for predictive biomarker discovery and validation (Tosolini et al., 2011).

B-lymphocytes, responsible for the production of highly specific antibodies, have also been shown to play some role in CRC. There is emerging evidence that increased density of a specific subtype of B-cells, CD20+ B-cells have been associated with increased OS, which could make them important for prognosis of CRC (Edin et al., 2019, Berntsson et al., 2016).

1.2.1.2 Innate immunity and colorectal cancer

Innate immunity has been described as the ‘first line of defence’ to cancer and other abnormal situations (Grizzi et al., 2013). There is evidence that the innate immune system may have antitumour effect that could impact prognosis (Ghiringhelli et al., 2009, Santos et al., 2014). The innate immune system may facilitate an antitumour response through the recognition of endogenous ligands released by dying cancer cells (Ghiringhelli et al., 2009, Tesniere et al., 2010, Vacchelli et al., 2015).

Macrophages, the primary source of pro-inflammatory cytokines, have been shown to play a key role in CRC. Cancer cells have been shown to ‘shape their interaction with cytokines’ by promotion of a distinctive balance of cytokines (described as M2-like) which encourage immunosuppression and cancer growth (Rigo et al., 2010). Additionally, tumour-associated macrophages (TAM) which target altered cells by secreting proteases that enhance invasion and metastasis, may be associated with a worse prognosis in CRC (Edin et al., 2013, Salama and Platell, 2008). Macrophages can also promote wnt-signalling, causing increased proliferation and expansion (Kaler et al., 2010). However, the exact role of macrophages in CRC is still a cause for debate.

The presence of other innate immune cells have been detected in CRC microenvironment including eosinophils, mast cells, natural killer (NK) cells and neutrophils (Grizzi et al., 2013). Increased levels of eosinophils surround a primary colorectal tumour may be a predictor of improved CRC outcomes (Harbaum et al., 2015). A similar association is seen in levels of mast cells with CRC stage and therefore they can be viewed as an independent prognostic factor for increased survival (Nielsen et al., 1999). The presence of NK cells can also be prognostic due to their cytotoxic effects when in contact with cancer-associated ligands (Malmberg et al., 2008, Corvaisier et al., 2005). Neutrophils are believed to account for around 15% of inflammatory cells and molecules associated with CRC, with this increasing in areas of tumour death (Salama and Platell, 2008). An elevated neutrophil/lymphocyte ratio has also been associated with decreased survival and increased rates of recurrence (Halazun et al., 2008, Jakubowska et al., 2020, Dell'Aquila et al., 2018).

The effect of the innate antitumour response facilitated by endogenous ligands has been reported to be most relevant in the terms of cancer cell death as the result of anthracyclines and oxaliplatin (Ghiringhelli et al., 2009, Tesniere et al., 2010, Vacchelli et al., 2015). Pattern Recognition Proteins (PRPs) which present endogenous ligands are an essential part of the innate immune response and therefore a focus for exploring the potential innate antitumour response (Takeuchi and Akira, 2010).

1.2.1.3 Pattern Recognition Proteins

PRPs, also referred to as Pattern Recognition Receptors (PRR), are expressed by cells of the innate immune system including macrophages and epithelial cells (Schroder and Tschopp, 2010). Downstream signalling cascades are the result of binding of two classes of molecules; pathogen-associated molecular patterns (PAMPs) – related to external pathogens and damage-associated molecular patterns (DAMPs) – associated with host cell damage of death (Ausubel, 2005, Rubartelli and Lotze, 2007). Mutations in genes encoding pattern recognition receptors linked to binding of molecules from both these categories (for example microbial antigens) alter innate immune activity in response to infection (Netea and van der Meer, 2011, Henckaerts et al., 2007). Defective recognition (caused by autosomal dominant or

recessive mutations) by three classes of PRPs including TLRs have been shown cause immune deficiencies resulting in increased infections from pathogens including Herpes simplex virus and *Staphylococcus Pseudomonas* (Netea et al., 2012). This would suggest that this downstream signalling cascade can activate pro-inflammatory, microbicidal and other non-specific immune responses required to contain abnormal molecules (Amarante-Mendes et al., 2018).

There are at least three classes of PRPs that have been shown to influence CRC phenotypes – Formyl Peptide Receptors (FPRs), Toll-like receptors (TLRs) and NOD-l-like receptors (NLRs) (Table 1.6). However, the focus of this thesis will be genes encoding three PRPs – *FPR1*, *TLR3* and *TLR4*. Single nuceotide polymorphisms (SNPs) from these 3 PRPs have been previously reported to be associated with oxaliplatin-based chemotherapy treatment outcomes including OS (Tesniere et al., 2010, Castro et al., 2011, Vacchelli et al., 2015).

FPR1 encodes a G protein-coupled receptor cell surface protein that binds to and is activated by molecules including N-Formylmethionine-containing oligopeptides (Migeotte et al., 2006). FPR1, which is expressed in mucosal tissues including the colon and rectum, recognises a variety of peptides including those released from damaged host cells (Li et al., 2016, Jeong and Bae, 2020). Upon activation by these molecules, the uncoupling of Gi-proteins occurs which activates downstream pathways associated with cell cycle progression and phagocytosis. Some of the proteins found in the signalling cascade are also influenced by the transmembrane protein EGFR, as shown in Figure 1.8 (Liu et al., 2012). *TLR3* and *TLR4* encode two transmembrane proteins which, when activated, influence a signalling cascade causing activation of genes associated with the activation of interferons and inflammation (Figure 1.8). The most notable difference between these two PRPs is their cellular localisation; TLR3 is localised to the endosome membrane, and TLR4 on the cell surface membrane (Kawasaki and Kawai, 2014). Genetic variation (including SNPs) within these genes have been shown to be associated with survival in CRC patients (Table 1.6).

Table 1.6. Pattern Recognition Receptor and reported colorectal cancer associations

PRR	Phenotype	Biological mechanism	Published study
Formyl Peptide Receptors			
FPR1	Increased expression associated with tumour invasion	Mechanism unknown	Li et al. (2017)
	Impact of OS for patients treated with oxaliplatin – explored as part of this thesis		Vacchelli et al. (2015)
FPR2	Increased expression associated with increased cancer progression	Promotion of epithelial-mesenchymal transition	Lu et al. (2019)
Toll like receptors			
TLR2	Decreased expression associated with tumour development*	Increased IL6 and associated pathways	Lowe et al. (2010)
TLR3/TLR4	Impact of OS for patients treated with oxaliplatin – explored as part of this thesis		Castro et al. (2011) Tesniere et al. (2010)
NOD-I-like receptors			
NOD1	Reduced expression in T-cells associated with decreased inflammation and associated tumorigenesis	Reduced IFNγ-mediated inflammation induced tumorigenesis	Zhan et al. (2016)
NOD2	Reduced expression associated with decreased inflammation and associated tumorigenesis	Induction of IRF4 causing MAPK and NK-κB inhibition	Udden et al. (2017)
NOD3	Reduced expression associated with cancer progression	PIK3 mediated inhibition of the mTOR signalling pathways	Ghiringhelli et al. (2009)
	Reduced expression increases cellular proliferation		
NOD4	Reduced expression correlates with cancer progression	Mechanisms unknown	Liu et al. (2015)
NOD5	Reduced expression is found in conjunction with inactivation of CD8+ T-cells	Dysregulation of MHC1 associated pathways	Yoshihama et al. (2016)
	Higher risk of cancer development		Ozcan et al. (2018) Huhn et al. (2018)

* Currently debated. Salcedo et al. (2010) found no colorectal cancer phenotype associated with reduced TLR2 expression

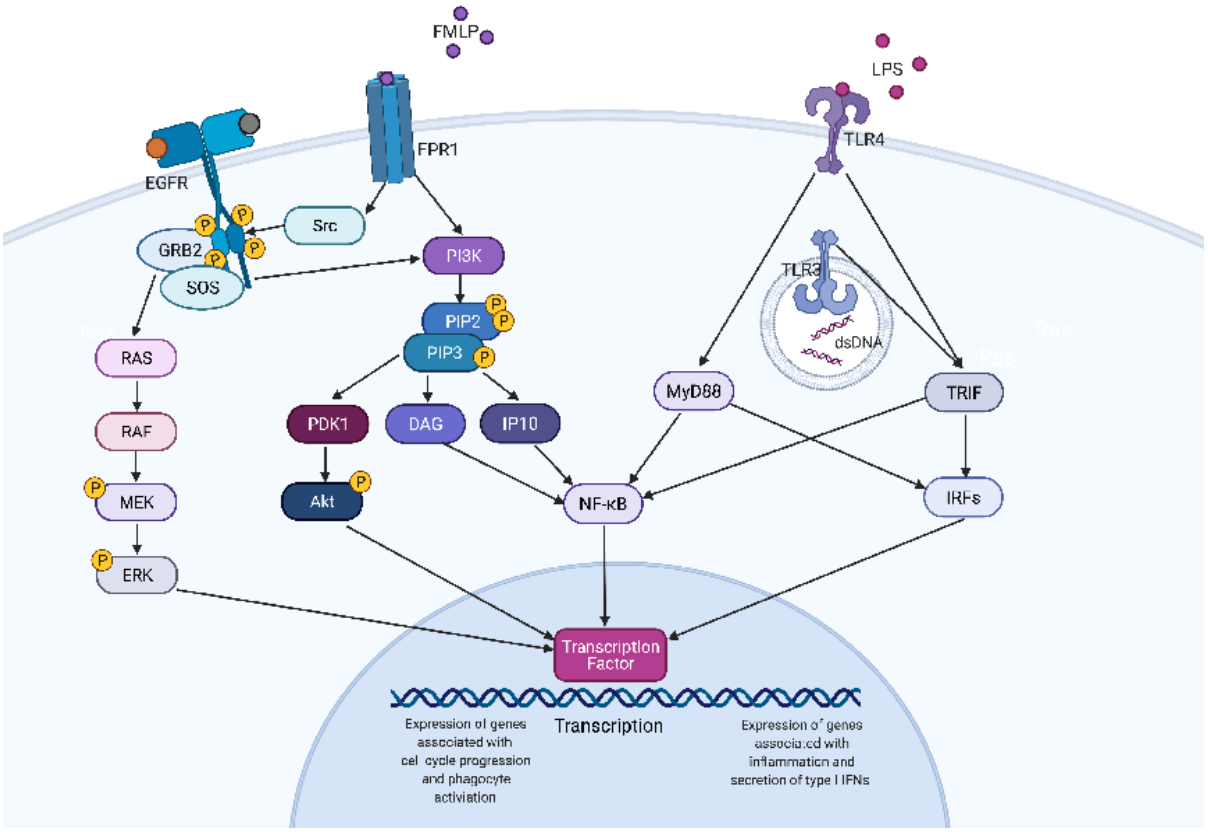

Figure 1.8. *FPR1*, *TLR3* and *TLR4* pathways to transcription
FPR1 is localised to the cell surface membranes and when activated, Gi-proteins uncouple, activating pathways associated with cell cycle progression and phagocytosis including pathways associated with cetuximab target EGFR. TLR3 is localised to the endosome membrane and TLR4 to the cell surface membrane. Dimer formation activates signalling by two adaptor proteins which begin a signalling cascade that results in activation of transcription factors and transcriptions of genes associated with inflammation and interferon action. Figure adapted from Liu et al. (2012) and Kawasaki and Kawai (2014).

1.3 Biomarkers for CRC outcomes

There are well established clinical factors associated with CRC progression, prognosis and treatment outcomes (Table 1.7). Clinicians may use information surrounding these in their decision-making about treatments for patients. For example, patients who are older are less likely to be given more aggressive treatments due their frailty (Millan et al., 2015). Other clinical factors shown to have a prognostic effect include sex, WHO performance status and white blood cell (WBC) count (Table 1.7). However, research has also highlighted the importance of CRC biomarkers (including acquired mutations and inherited variation in germline DNA) in the era of personalised medicine.

1.3.1 Somatic biomarkers

Traditionally the search for biomarkers associated with CRC outcomes has largely been associated with somatic mutations in tumour tissues (Section 1.1.3.2, Table 1.1 and Table 1.8), with some having clinical implications.

Most notably, *KRAS* and *NRAS* [*RAS*] mutations have been shown to have both predictive and prognostic effects due to their involvement with cell proliferation. The most common mutations in *RAS* genes occur at exon 2 (codons 12 and 13). Due to the genes involvement in the EGFR pathway, *RAS* mutations are a predictive biomarker for lack of response to anti-EGFR therapies cetuximab and panitumumab (Chung et al., 2005, Cutsem et al., 2010, Gong et al., 2016). Due to this lack of response, *RAS* mutations are not in the inclusion criteria for treatment of aCRC with cetuximab or panitumumab (NHS National Institute for Health and Clinical Excellence (NICE), 2017, Van Cutsem et al., 2014)

Despite other tumour biomarkers including *BRAF* mutations being found to have a negative effect on survival, their use as predictive biomarkers for specific CRC treatments is limited (Sanz-Garcia et al., 2017). However, through discovery of new CRC treatments such as Lonsurf, better outcomes for patients with *BRAF* mutant tumours are possible (van der Velden et al., 2017).

Table 1.7. Clinical factors associated with CRC prognosis and treatment outcomes

Factors	Published study	Analysed outcome	Patients recruited	Outcomes reported
Age at diagnosis	Brenner et al. (2014)	Epidemiology	N/A	↑ risk, older age
	van Eeghen et al. (2015)	Prognosis	621	↓ OS, older age (HR=1.02, *P*<0.05)
AJCC stage at diagnosis	Amin et al. (2017)	Prognosis	N/A	Worse prognosis with each stage
	Miller et al. (2019)	Prognosis	N/A	↓ 5-year survival (91% stage II vs 12% for IV)
ALKP levels	Köhne et al. (2002)	TO for 5-FU	3,825	↓ survival, high ALKP (12.8 vs 13.5 months, *P*<0.01)
	Hung et al. (2017)	Prognosis	10,800	↓ 5 year OS rate, high ALKP (Colon 72% vs 78%, P<0.01; Rectum 65% vs 72% *P*<0.01)
Number of metastatic sites	Köhne et al. (2002)	TO for 5-FU	3,825	↓ survival, 2+ sites (12.0 vs 13.6 months, *P*<0.01)
	Wang et al. (2020)	Prognosis	26,170	↓ prognosis
Obstruction and perforation at presentation	Steinberg et al. (1986)	Prognosis	527	↓ DFS (Obstruction RR=1.40, P=0.04; Perforation RR=3.40 *P*<0.01)
Primary tumour resection status	Faron et al. (2015)	Prognosis	1,155	↑ OS (HR=0.63, *P*<0.01)
	Wang et al. (2016)	TO for bevacizumab	199	↑ OS (22.5 vs 17.8 months, *P*<0.01)
Preoperative platelet count	(Wan et al., 2013)	Prognosis	1,513	↓ survival, high count (HR=1.66, *P*<0.01)
Sex	Majek et al. (2013)	Prognosis	164,996	↑ 5-year survival, in women (65% vs 62%, *P*<0.01)
	White et al. (2018)	Prognosis	N/A	↑ 5-year survival, women (Stage II) (87% vs 82%, *P* not reported)
	Schmuck et al. (2020)	Prognosis	185,967	↑ OS, women (HR=0.80, *P*<0.01)

Tumour sidedness	O'Dwyer et al. (2001)	To for 5-FU	1,120	↓ OS, right-sided (10.9 months vs 15.8 left-sided)
	Kamran et al. (2018)	Prognosis	367	↓ 5 year OS, right-sided (24% vs 46%, *P*<0.01)

AJCC=American Joint Committee on Cancer, ALKP=alkaline phosphatase, FU=fluorouracil, N/A=not applicable. OS=overall survival, TO=treatment outcome

Table 1.8. Somatic variants associated with CRC prognosis and treatment outcomes

Variant	Published study	Analysed Outcome	Patients Recruited	Outcomes Reported
BRAF mutation	Di Nicolantonio et al. (2008)	Response to panitumumab or cetuximab	113	↓ Response (OR not reported, P=0.03)
	Richman et al. (2009)	Prognosis	711	↓ OS (HR=1.82, *P*<0.01)
	Tran et al. (2011)	Prognosis	524	↓ OS (median 24.3 months less, P<0.01)
	Guan et al. (2020)	Prognosis	74	↓ OS V600E vs all other *BRAF* mutations (HR=2.94, *P*=0.02)
CpG Island Methylator Phenotype (CIMP)	Samowitz et al. (2005)	Prognosis	886	↓ OS high CIMP (median 4.7 months less, *P*<0.01)
	Barault et al. (2008)	Prognosis	582	↓ OS low CIMP (HR=1.86, *P*<0.01) ↓ OS high CIMP (HR=2.90, *P*<0.01)
	Ogino et al. (2009a)	Prognosis	649	↓ CSS high CIMP (HR=0.41, *P* not reported)
	Ouchi et al. (2015)	TO for anti-EGFR therapies		↓ ORR high CIMP (35.7% WT vs 6.3% mut, *P*=0.03)
	Kim et al. (2017)	Prognosis	157	↓ DFS CIMP high (HR=2.01, *P*=0.02)
CIN	Walther et al. (2008)	Prognosis	10,126	↓ OS (HR=1.45, *P*<0.01)
KRAS mutation	Andreyev et al. (1998)	Prognosis	2,721	↓ OS (HR=1.25, *P*<0.01)
	Andreyev et al. (2001)	Prognosis	4,268	↓ FFS (codon 12 mutation) (HR=1.3, *P*<0.01) ↓ OS (codon 12 mutation) (HR=1.29, *P*<0.01)
	Lièvre et al. (2006)	Response to cetuximab	30	↓ Response (OR not reported, *P*<0.01) ↓ OS

				(Median 9.4 months less, *P*=0.02)
	Karapetis et al. (2008)	Cetuximab benefit	572	↓ ORR (12.8% WT vs 1.2% mut) ↓ effect of cetuximab on OS ($P_{Interaction}$=0.01)
	Eklöf et al. (2013)	Prognosis	414 (CRUMS cohort)	↓ CSS (HR=1.48, *P*<0.01)
MSI/MMR (early stage)	Popat et al. (2005)	Prognosis	7,642	↑ OS (HR=0.65, heterogeneity *P*=0.16)
	Bertagnolli et al. (2009)	TO for FU/LV and IFL	1,264	↑ DFS IFL (HR=0.76, *P*=0.03)
	Hutchins et al. (2011)	TO for FU and FA	1,913	↓ RecR (RR=0.53, *P*<0.01)
	Lochhead et al. (2013)	Prognosis	1,253	↑ CSS (*BRAF* WT HR=0.26, *P*<0.01; *BRAF* mut HR=0.44, P<0.01)
MSI/MMR (late stage)	Tran et al. (2011)	Prognosis	524	↓ OS (median 10 months less, *P*<0.01)
	Smith et al. (2013a)	Prognosis	1,976	↓ OS (HR=1.60, *P*<0.01) ↓ PFS (HR=1.66, *P*<0.01)
PIK3CA mutation	Ogino et al. (2009b)	Prognosis	82	↓ CSS (HR=2.23, *P* not reported)
NRAS mutation	Schirripa et al. (2015)	Prognosis	786	↓ OS (HR=1.91, *P*<0.01)
RAS [KRAS and *NRAS]* mutation	Khattak et al. (2015)	TO for anti-EGFR therapies	2,014	↑ OS (*RAS* wt vs anti-VEGF) (HR=0.77, *P*=0.02) ↑ ORR (*RAS* wt vs anti-VEGF) (OR=1.31, *P*<0.01)
	Guren et al. (2017)	TO for cetuximab	223	↓ OS (20.3 months vs 24.6 months, *P*=0.03)
Reduced heterozygosity at 18q	Jen et al. (1994)	Prognosis	145	↓ OS (HR=2.83, *P*<0.01)

	Ogunbiyi et al. (1998)	Prognosis	151	↓ DFS (RR=1.65, *P*=0.01)
SMAD4 (protein and mRNA levels)	Alazzouzi et al. (2005)	Prognosis	86	↓ OS (↓ expression) (*P*=0.02) ↓ PFS (↓ expression) (*P*=0.02)
		TO for 5-FU		↓ OS (↓ expression) (1.4 years vs >9.3 years, *P*<0.01)
TP53 mutation	Russo et al. (2005)	Prognosis	3,583	↓ OS (distal colon) (RR=2.52, *P*=0.01) ↓ OS (proximal colon)* (RR=1.36, *P*=0.03)
		Prognosis	18,766	↑ risk of death (RR 1.32, *P* not reported)

AJCC=American Joint Committee on Cancer, CSS=cancer-specific survival, DFS=disease-free survival, EGFR=epidermal growth factor receptor, FA=folinic acid, FFS=failure-free survival, FU=Flurouracil, IFL=irinotecan, FU, and LV, LV=leucovorin, mut=mutant, ORR=overall response rate, OS=overall survival, RecR=recurrence rate, RR=risk ratio, TO=treatment outcome, WT=wild-type, *=trend towards significance (did not meet significance threshold when adjusted for multiple tests).

1.3.2 Germline biomarkers

There is now an increasing amount of research being performed into the role of germline variants in cancer risk, progression and clinical outcomes. The approaches for these investigations have predominantly focused on candidate variants, including genes found within pathways that CRC therapeutics target (Smith et al., 2015). The results of these investigations (along with some which search the whole genome) are germline variants which have associations with CRC outcomes moderate effect sizes (Table 1.9).

Molecular interactions of CRC and treatments may not be fully understood and may result in shortcomings of this method (Mármol et al., 2017). This, coupled with the fact that the already established mechanisms for progression may not be associated with clinical outcomes, contribute to the lack of replication of positive findings and limited clinically applicable variants discovered using this approach (Phipps et al., 2016, Pasche and Yi, 2010).

Table 1.9. Germline variants significantly associated with CRC prognosis and treatment outcomes

Variant(s)	Published study	Analysed outcome	Patients recruited	Outcomes reported
rs10817938	Hu et al. (2019)	TO for oxaliplatin	580	↑ OS (OR=0.73, *P*<0.01)
rs1801133	Custodio et al. (2014)	TO for oxaliplatin	202	↑ Recurrence risk (RR=3.57, *P*=0.02)
rs2072493	Klimosch et al. (2013)	Prognosis	613	↓ CSC (HR=1.89, *P*<0.01) ↓ OS (HR=1.60, *P*=0.01)
rs2231142	(Hu et al., 2019)	Prognosis	580	↓ DFS (HR=0.68, *P*<0.01) ↓ DFS (HR=0.67, *P*<0.01)
rs209489	Phipps et al. (2016)	Prognosis	3,494	↓ OS (HR=1.80, *P*<0.01)
rs3775291	Castro et al. (2011)	TO for oxaliplatin	565	↓ CSC (HR=2.00, *P*=0.03)
rs4986790	Tesniere et al. (2010)	TO for oxaliplatin	338	↓ OS (HR=1.40, *P*=0.05) ↓ PFS (HR=1.40, *P*<0.05)
rs5030740	Li et al. (2019b)	TO for oxaliplatin	166	↓ PFS (HR=1.86, *P*<0.01)
rs867228	Vacchelli et al. (2015)	TO for oxaliplatin	311	↓ OS (HR=2.10, *P*<0.01) ↓ PFS (HR=1.90, *P*<0.03)
rs9929218	Smith et al. (2015)	Prognosis	7,635	↓ OS (HR=1.14, *P*<0.01)

CSC=colorectal specific survival, DFS=disease-free survival, OS=overall survival, RR=risk ratio, TO=treatment outcome

1.4 GWAS

An alternative approach of search for germline variants (such as SNPs) is now being employed – Genome Wide Association Studies (GWAS). The principle of GWASs allows for a comprehensive and unbiased scan of the entire genome for statistical associations with different phenotypes without the need for biological knowledge of mechanisms or candidate variants (Hirschhorn and Daly, 2005, McCarthy et al., 2008). The success of GWASs relies on genetic information from genome mapping projects including the International HapMap project and 1,000 genomes projects (Gibbs et al., 2003, Frazer et al., 2007, Auton et al., 2015). As a result of these projects, a more detailed understanding of previously unknown variants can be used to determine statistical associations with cancer outcomes (Table 1.9).

GWASs have been successfully used to discover germline variants associated with a variety of diseases with known genetic components, including neurological diseases like schizophrenia and other disorders including type I and II diabetes (Bergen and Petryshen, 2012, Pociot, 2017, Xue et al., 2018). There have also been variants significantly associated with the risk of developing a number of solid tumours including breast cancer, gastric cancer, lung cancer and CRC (Easton et al., 2007, Ferreira et al., 2019, Tanikawa et al., 2018, Bossé and Amos, 2018). As of 2019, 79 loci associated with CRC susceptibility have been identified as the result of GWASs (Law et al., 2019). There have however, been limited use of this methodology on CRC outcomes, with those performed largely focusing on survival (Phipps et al., 2016, Penney et al., 2020, Summers, 2019).

1.4.1 The ‘common disease, common variant’ hypothesis

The detection of variants by GWAS is based on principles described in the ‘common disease, common variant’ (CD/CV) hypothesis (Figure 1.9) (Reich and Lander, 2001). The hypothesis states that common inherited disorders are likely to be as the result by common genetic variants in the population (Wang et al., 2005). There is an inverse correlation between allele frequency and the prevalence of the disease in population (Bush and Moore, 2012). Therefore, common variants cannot have large effect sizes

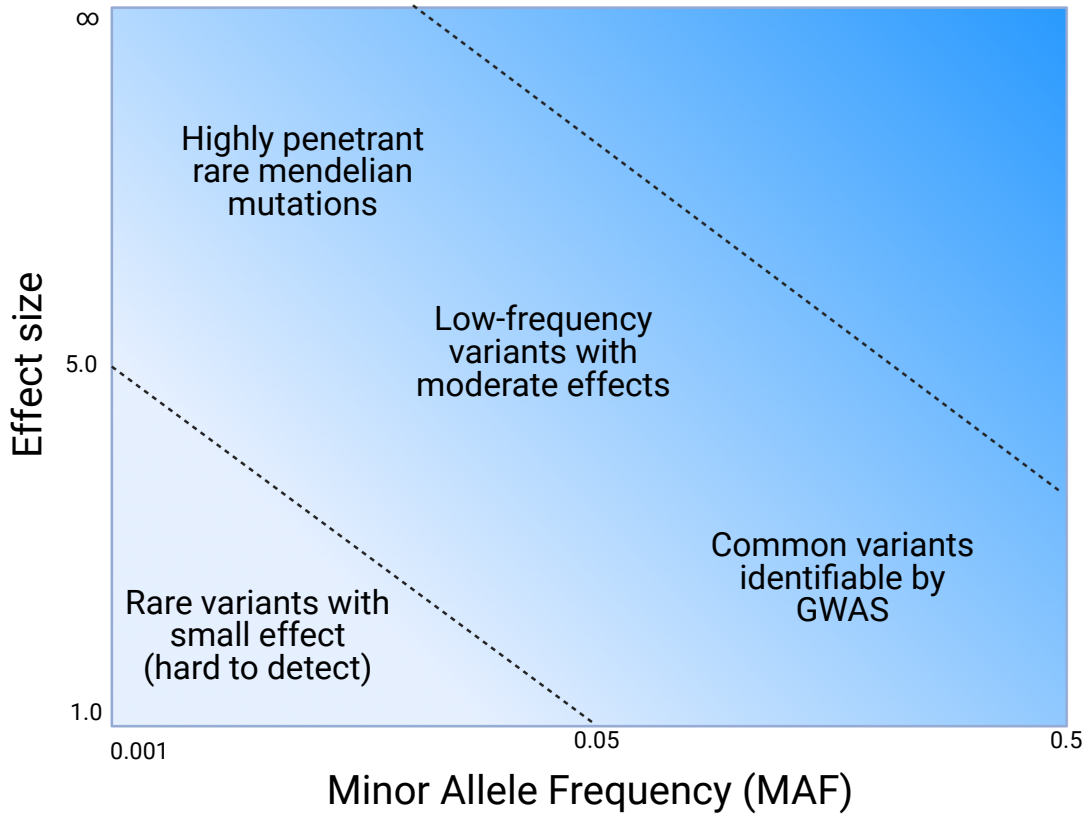

Figure 1.9. The allelic spectrum of human disease predisposition
Adapted from McCarthy et al. (2008). Alleles with high penetrance for Mendelian disorders are extremely rare and have large effect sizes (top left), while the majority of findings from GWASs are associations of common SNPs with small effect sizes (bottom right). Most interest and emphasis in identifying associations with phenotypes lie between the two dotted lines.

on common diseases as the effect size of these variants must be small relative to those found in rarer disorders. Additionally, if alleles have moderate effect sizes and common disorders have an element of heritability (as shown in the genetic mechanisms of cancer predisposition), common disease susceptibility is likely due to the contributory effects of multiple common and unlinked variants. This would mean affected individuals who are unrelated would share a significant amount of disease associated alleles (Wang et al., 2005). Therefore effect sizes and allelic frequencies of statistically significant variants detected directly influence statistical power of studies based on sample size considerations (Wang et al., 2005).

1.4.2 Single Nucleotide Polymorphisms

Single Nucleotide Polymorphisms (SNPs) are sometimes described as the simplest form of DNA variation in individuals (Shastry, 2009). SNPs are single base-pair changes in a DNA sequence with minor allele frequency (MAF) of >1% in the general population and have a unique identifier (rs number) with a discrete genomic location (Risch, 2000, Sherry et al., 2001). These largely take the form of base substitutions and insertions or deletions (indels) (Chen et al., 2009b). These can either affect the codons for amino acids through missense, nonsense or frameshift mutations or have no effect (termed ‘silent mutations’). Estimations of average SNP density vary, however the distribution has been shown to be different across regions or the same chromosome and between whole chromosomes (Chen et al., 2009b, Ke et al., 2004, Chanock, 2001). There is evidence that this variation is due to the selection pressure of natural selection (Zhao et al., 2003). This means that the effect of most SNPs are silent – with no impact on protein structure (Erichsen and Chanock, 2004). However, SNPs can have other functional consequences that do not directly influence protein coding including changes transcription factor (TF) binding sites, mRNA stability and influence gene expression through the process of expression quantitative trait loci (eQTLs) (Durbin et al., 2010, Griffith et al., 2008).

1.4.3 Linkage Disequilibrium (LD)

The variation in SNP density which is due to selection pressure is also seen in the principle of LD, the association between loci on the same chromosome which is non-random and variable both across genomes and between different populations (Bush and Moore, 2012, Hirschhorn and Daly, 2005). LD is based on the principle of linkage, where areas of the same chromosome are inherited together through generations of recombination events (Slatkin, 2008).

LD can be measured by 2 statistics; D' – most common in population genetics – and r^2 – used in association studies (Hill and Robertson, 1968). D' (between -1 and 1) is derived from D (coefficient of disequilibrium, measure of linkage between two variants) and the theoretical maximum difference between the observed and expected haplotypes (Lewontin, 1964). r^2 (measured between 0 and 1) is a measure of correlation, inversely proportional to the sample size, and is often used to measure the linkage between two loci (VanLiere and Rosenberg, 2008). The higher the r^2 value, the more likely that a change in the first variant is observed, and is also present in the second variant, meaning that the variants (in most cases SNPs) are in LD. A high measure of LD is only possible when the alleles of the SNPs are correlated, are found on the same ancestral haplotype and have a similar MAF (Takeuchi et al., 2005). This means that genotyping of only one of these variants is required which prevents redundancy of information and can optimise the process of GWASs (Bush and Moore, 2012). It has been estimated that a specific subset of <1 million SNPs from the International HapMap Project can cover >80% of commonly occurring SNPs in European populations (Li et al., 2008). The genotype information of these SNPs is acquired using chip-based arrays which allow for cost effective genotyping and imputation of variants that are not directly genotyped (Li et al., 2009).

The presence of SNPs in LD can have two different implications on variants detected as the result of a GWAS. Either the SNP most statistically associated with the phenotype is directly influencing the biological mechanism causing this or the causal SNP has not been genotyped and is found in linkage with genotyped SNP (MacArthur et al., 2014). This means that it cannot always be assumed that the discovered SNP

is causal and in some instances, the process of fine-mapping is used to determine if this is the case (Raychaudhuri, 2011).

1.4.4 Genotyping and imputation

The process of chip-genotyping is a cost effective way to genotype a large number of single variants required for large coverage of the genome that underpins the process of a GWAS.

It is now common practice not to directly genotype all variants investigated through association studies. A large amount of genetic information comes from the process of genomic imputation. Genomic imputation predicts the genotypes using a full sequenced reference panel such as HapMap or 1,000 genome project (Li et al., 2009). This is done through the process of estimating the missing haplotype (known as haplotype phasing). Due to the level of uncertainty in the process, a probability for each possible genotype at each location is produced. In some instances, this is then recoded to a specific allele, rather than the probability originally reported (Li et al., 2009). These *in silico* genotypes greatly increase the genome coverage of SNPs available for testing for association. The process of imputation can increase the impact of GWASs through increasing the statistical power of initial GWAS, allowing for fine-mapping to discover causal variants and enabling meta-analyses of multiple studies together (Marchini and Howie, 2010).

A commonly used method for imputation is using the software IMPUTEv2 which reports genotypes as an infoscore (Howie et al., 2009). Infoscores (a value between 0 and 1) are an indication of the quality of imputation. The closer the score is to 1, the more likely it is that the imputed genotype reflects the actual genotype of the SNP (Huang et al., 2015). This means that the filtering of SNPs by infoscore during quality control (QC) steps before analyses is often performed to ensure poorly imputed SNPs are not analysed. The threshold for this varies, with an infoscore of >0.4 often being considered an appropriate threshold but most investigations use the highly stringent threshold of >0.8 (Zheng et al., 2015, Huang et al., 2015). Historically, genotyping and imputation technologies had better coverage for SNPs with a MAF>5% (Barrett and

Cardon, 2006, Flannick et al., 2012). Whilst there are some technologies with the ability to detect low frequency variants, the coverage of traditional chip-genotyping, coupled with the reported increased statistical power at MAF>5%, has resulted in the threshold of 5% being accepted for common variants in GWASs (Fan et al., 2011, Fadista et al., 2016).

1.4.5 Visualising and Interpreting GWAS results

Due to the nature of GWAS, there are specific methodologies and considerations required to gain the full insight from the dataset – from statistical to biological impacts of the results. This includes visualisation and downstream analysis of the results from the primary analyses.

1.4.5.1 Sample size and power

Due to the previously discussed CD/CV hypothesis (Section 1.3.1), it has been established that variants that contribute to complex inherited traits are likely to have small effect sizes (Reich and Lander, 2001). This would mean that a large sample size is required for a GWAS to have the ability to detect any statistically significant associations (Hirschhorn and Daly, 2005). However, in contrast to some statistical investigations there are multiple statistical tests occurring at once which requires an adjustment to the significance threshold to account for this multiple testing. This adjustment prevents the likelihood of false positive results (Forstmeier et al., 2017). The *de facto* adjusted significance threshold for genome-wide significance is $P<5.0\times10^{-8}$ (Jannot et al., 2015). This is as the result of a Bonferroni correction of standard significance (P<0.05) for 1,000,000 independent SNPs (Risch and Merikangas, 1996, Bland and Altman, 1995). More recently a second threshold of suggestive significance (*P*<1.0x10-5) has been established to highlight loci with a potential association that requires variation; however, most power considerations are still performed based on genome-wide significance.

Sample size is linked to statistical power of the study. Power is defined as 1-β (where β is the probability of Type II errors), the probability of correctly rejecting the null

hypothesis. Some of the considerations required for power calculations are outside the control of trial or investigation design, including the effect sizes. MAF also plays a large role in power calculations but SNPs with smaller MAF can be filtered out during the QC process to reduce the burden of multiple testing. Power can also be maximised by an investigator selection of individuals, sample size, and additional QC steps (Sham and Purcell, 2014). The power of GWASs varies between studies but it has been established that the variants that reach genome-wide significance in a higher powered study are more likely to represent genuine results than those found in lower powered studies (Sham and Purcell, 2014).

1.4.5.2 Manhattan Plots

Visualisation of GWAS results is commonly done using a Manhattan Plot (Figure 1.10A). Manhattan plots plot the chromosomal position on the x-axis and the association for each SNP (-log10(*P*)) on the y-axis (Turner, 2014). Red and blue lines are commonly placed on the plot indicate the two different significance thresholds – genome-wide ($P<5.0\times10^{-8}$) and suggestive ($P<1.0\times10^{-5}$) significance respectively (Figure 1.10A).

1.4.5.3 Quantile-Quantile (Q-Q) plots

Q-Q plots (Figure 1.10B) are utilised to determine if there are any issues with the data including relatedness and presence of population stratification. The plot shows the expected distribution of *P*-values if the null hypothesis is correct on the x-axis and the observed *P*-values on the y-axis. SNPs that show a significant association will deviate from the straight diagonal line on the plot. However, a complete deviation from the line would indicate issues in population stratification. This is also measured using the lambda (λ) statistic. The closer to 1, the more likely the appropriate adjustment for population substructure has been applied (Reed et al., 2015).

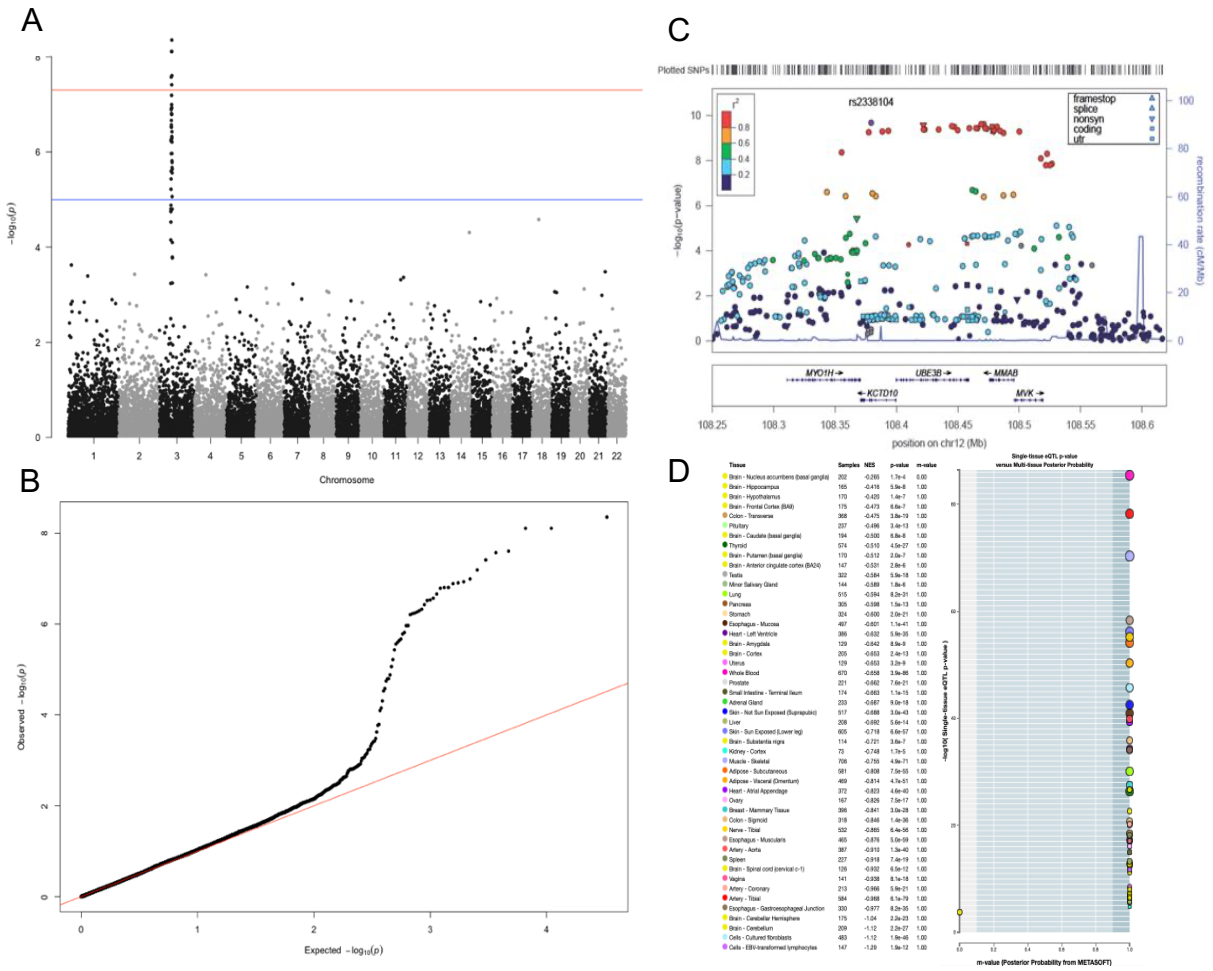

Figure 1.10. Data visualisation for results of a GWAS

(A) Manhattan plot showing the significance (-$\log_{10}$(*P*), Y-axis) of tested SNPs against genomic location (Chromosomal position, X-axis). Genome wide significance (red line)=$P<5.0\times10^{-8}$. Suggestive association (blue line)=$P<1.0\times10^{-5}$. Figure adapted from Turner (2014) and plotted using the qqman package in R. (B) Q-Q plot showing expected vs observed P-values for SNPs tested in (A). Upward deviation from the red line could indicate genomic inflation and therefore artificially high peaks in a Manhattan plot. (C) Regional association plot. Plot made using LocusZoom and adapted from Pruim et al. (2010). $-\log_{10}$(P) (y axes) of the SNPs are shown according to their chromosomal position (x axes). The sentinel SNP (purple) is labelled. The colour intensity of each symbol reflects the extent of linkage disequilibrium (LD) with the sentinel SNP, deep blue (r^2=0) through to dark red (r^2=1.0). Genetic recombination rates, estimated using 1,000 Genomes Project samples, are shown with a blue line. Physical positions are based on NCBI build 38 of the human genome. Also shown are the relative positions of genes and transcripts (direction of transcription indicated by the arrow) mapping to the region of association. Genes may not be to scale as they are drawn to show their relative positions. (D) Example multi-tissue eQTL plot (rs1805008**).** Plot created using GTEx online portal.

Both Manhattan Q-Q plots can be created using the qqman package in R, as shown in Figures 1.10A and B (Turner, 2014).

1.4.5.4 Regional association analyses

Once significant results of a GWAS have been established, visual assessment of regions of the genome in which the loci are found can be performed. This allows for identification of genes and regulatory elements such as microRNA transcription sites which may be in the same haplotype as the significant loci. LocusZoom is a web-based tool for plotting the regions associated with GWAS results based on a pre-specified area of the genome (Pruim et al., 2010). LocusZoom uses the UCSC Genome Browser to map the SNPs of interest and uses LD information based on results from HapMap Phase II to give an overview of the nearby genes and recombination events (Frazer et al., 2007, Pruim et al., 2010).

1.4.5.5 Expression quantitative trait loci (eQTL) analyses

Additional analyses can be performed to put the significant loci into context in relation to biological function. eQTLs take into account the role of a variant in influencing the expression phenotype of other genes (Nica and Dermitzakis, 2013). There are two types of eQTL – those close to the variant (cis-eQTLs) and those further from the variant (trans-eQTLs) (Battle et al., 2017). Information on these eQTLs can be accessed from the Genotype-Tissue Expression (GTEx) project, which is an open access database where the results of individual SNPs can be searched (The GTEx Consortium, 2013, Carithers and Moore, 2015). These results can be visualised in a multi-tissue eQTL plot (Figure 1.10D).

The GTEx database has been used by a number of studies exploring CRC risk variants and candidate susceptibility genes (Loo et al., 2012, Hulur et al., 2015, Closa et al., 2014). Exploration of eQTLs has also been performed for a number of other cancers including breast cancer and lung cancer (Beesley et al., 2020, Fan et al., 2019).

1.4.5.6 Gene-based and gene-set analysis

In some instances, more biological or clinical meaning can be gleaned from the results of a GWAS by analysing multiple SNPs at once in order to determine their additive effect. This can be in terms of the effect across a single gene or through how genes interact into pathways (de Leeuw et al., 2015). The gene-based and gene-set analyses allow for significant loci to be placed into a wider biological context that may help understand how multiple significant loci interact with each other.

Analysis of genes and gene-sets is performed using MAGMA (de Leeuw et al., 2015). The process behind this consists of two main steps, an annotation step based on a reference panel and genomic built (such as the 1,000 genomes panel and GRCh37/hg19) and the gene-level analyses step. Gene-level analyses are performed using a methodology based upon multiple linear principal components regression using the GWAS summary statistics (de Leeuw et al., 2015). These, like with GWAS results, need to be corrected for multiple testing – using the Bonferroni correction methodology in this case – for the around 20,000 genes identified in the human genome (Bland and Altman, 1995, Salzberg, 2018). Gene-set (or pathway) analyses are derived from the results of gene based analyses. Gene-set analyses are corrected for multiple testing based on the false discovery principle and produce an adjusted significance statistic with a threshold of 0.05 (de Leeuw et al., 2015, Forstmeier et al., 2017).

1.4.5.7 The PubMed database

To gain insight of the biological relevance of results of a GWAS and the downstream investigation including gene-based and gene-set analyses, further searches for appropriate scientific publication is required. After identification of variants, a search through the PubMed database is performed to gain further understanding of the clinical relevance of the findings. This database is facilitated by the National Centre for Biotechnology Information (NCBI) and contains more than 30 million biomedical publications from life science journals, MEDLINE and books (Macleod, 2002).

1.5 Hypothesis and aims

The main hypothesis of the thesis is that there are, yet to be discovered, novel, common variants that are biomarkers for treatment outcomes in patients with aCRC. This can be broken down into these specific aims:

- Investigate the role of common germline variants (MAF>5%) in response to oxaliplatin-based chemotherapy in the COIN and COIN-B datasets.
    - Investigate the role of variants of specific PRPs – *FPR1*, *TLR3* and *TLR4* – on oxaliplatin outcomes including OS and response.
    - Perform a GWAS to identify common variants significantly associated with response to chemotherapy.
- Investigate the genetic and clinicopathological factors that contribute to treatment outcomes of cetuximab in patients from COIN and COIN-B trials.
    - Determine the genetic and clinicopathological factors that influence response to cetuximab by performing exploratory factor analyses.
    - Perform a GWAS on a subset of individuals who are more likely to respond to cetuximab to identify common variants significantly associated with this phenotype.

2. Chapter 2: Materials and Methods

2.1 Resources used in this thesis

2.1.1 Hardware

Analyses were performed using an Apple (Cupertino, USA) MacBook Pro (Retina, 15.4 inch, 3.1 GHz Quad-Core Intel Core i7, 16 GB 2133 MHz LPDDR3), using operating system MacOS Big Sur. Advanced Research Computing at Cardiff (ARCCA) granted access to Cardiff University's high-performance cluster (HPC) Hawk (earlier Raven) which was used via command line-based remote access for processes that required intensive computation such as file conversions.

2.1.2 Software

Analysis was performed using software programs designed for statistical and genetic analyses. The majority of this was performed using R-version 3.5.2 (CRAN Corporation, Vienna, Austria), an open-source language downloaded from http://www.r-project.org, R Core Team (2018). This was used in conjunction with RStudio version 1.2.5033 (Orange Blossom release, RStudio, Inc., Boston, MA), an integrated development environment (IDE) downloaded from https://www.rstudio.com (RStudio Team, 2019). Other statistical analyses were performed using STATA/SE version 16.1 (StataCorpLP, College Station, TX).

Four additional software programs designed for genetic analysis were utilised. GTOOL (Genomics Software Suite, University of Oxford) was used to convert genotype files for downstream analyses and was downloaded from https://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html. SNPTEST version 2 (Marchini and Howie, Oxford, UK) was used for calculating scores and was downloaded from https://www.well.ox.ac.uk/~gav/snptest/. PLINK versions 1.9 and 2.0 (Shaun Purcell, Harvard, MA) were used for performing GWASs and downloaded

from http://pngu.mgh.harvard.edu/purcell/plink/ and MAGMA versions 1.07 and 1.08 (CTG Lab, Amsterdam, Netherlands) were used for gene-based and gene-set analyses and downloaded from https://ctg.cncr.nl/software/magma (Purcell et al., 2007, Chang et al., 2015, de Leeuw et al., 2015).

2.1.3 Packages and modules

Packages for R and STATA are downloaded and installed within the software environment from the Comprehensive R Archive Network (CRAN) and STATA repositories respectively. Packages and modules used for this thesis are listed in Table 2.1.

2.1.4 Web links

Web-based packages were also used for additional analyses. The Genotype-Tissue Expression project database (GTEx), a public resource for gene expression data can be found at https://gtexportal.org/home (Carithers and Moore, 2015). LocusZoom, a suite of tools to provide fast visualisation of GWAS results for research and publication can be accessed at http://locuszoom.org (Pruim et al., 2010).

2.2 My contribution and others contributions

The methods and resources developed in this thesis are a combination of my work and that of others prior to the start of this project (Figure 2.1). Sample collections, genotyping and some quality control (QC) was performed prior to commencement of this project. All other analyses were performed by myself unless stated otherwise.

2.3 Patients and samples

Two thousand, six hundred and seventy-one patients with locally advanced or metastatic colorectal adenocarcinoma who had not previously received chemotherapy for advanced disease were recruited into the MRC clinical trials COIN

Table 2.1. Packages and modules used in this thesis

Package	Software	Purpose	Reference
Base	R	Functions automatically loaded into R. Basic database functions	R Core Team (2018)
BiocManager	R	Repository of Bioconductor project. Used to download other packages including qvalue	Morgan (2019)
Car	R	Companion to applied regression. Recode functionality	Fox and Sanford (2019)
devtools	R	Development tool for R packages. Required to download other packages such as genpwr	Wickham et al. (2020b)
Dplyr	R	Database manipulation	Wickham et al. (2020a)
GenABEL	R	SNP association analyses and associated data handling.	GenABEL project developers (2013), Karssen et al. (2016)
Genpwr	R	Power calculations for genetic models	Moore and Jacobson (2020)
IPDMETAN	STATA	Exploratory factor analyses	Fisher (2014)
ggplot2	R	Data visualisation	Wickham (2016)
Metan	STATA	Metanalyses	Harris et al. (2008)
Purr	R	Functional programming	Henry and Wickham (2020)
qqman	R	Q-Q and Manhattan plots for visualising GWAS data	Turner (2014)
Qvalue	R	Calculation of false discovery rate statistics	Storey et al. (2020)
RColorBrewer	R	Colour pallets for visualisation of data	Neuwirth (2014)
Stats	R	Functions automatically loaded into R. Basic database functions. Calculates chi-squared tests.	R Core Team (2018)
survival	R	Survival analyses	Therneau (2020)
survminer	R	Survival analyses including data visualization	Kassambara et al. (2021)
survSNP	R	Power calculations for survival analyses	Owzar et al. (2012)
Utils	R	R utility functions	R Core Team (2018)

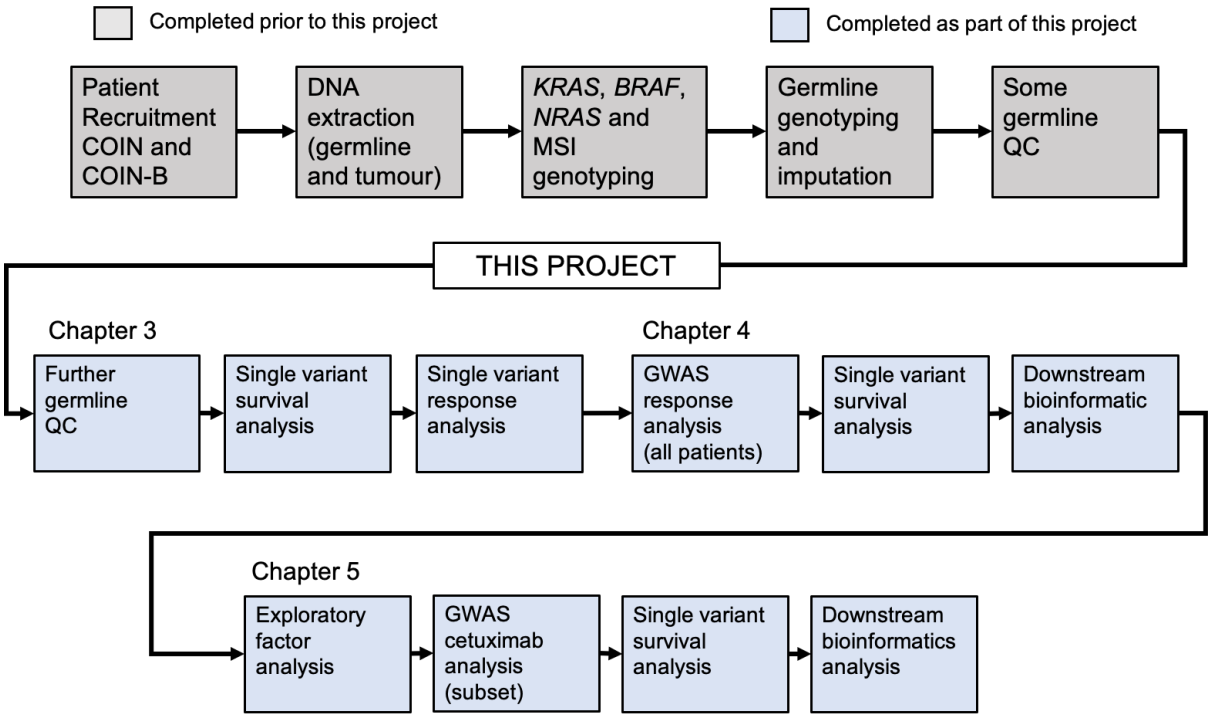

Figure 2.1. Workflow of processes performed before and during this project
Processes completed prior to the start of the project are highlighted in grey and processes completed as part of this project are highlighted in blue. Patient recruitment was performed by the Medical Research Council's (MRC) Clinical Trials Unit. DNA extraction and tumour DNA analysis were performed at Cardiff University. Whole genome SNP genotyping was performed at the King Faisal Specialist Hospital and Research Centre, Saudi Arabia. Initial QC of whole genome genotyping data was performed by Richard Houlston's group at the Institute of Cancer Research.(ISRCTN27286448) and COIN-B (ISRCTN3837568; Chapter 1, Sections 1.4.4 and 1.4.5). COIN patients were randomised 1:1:1 to receive continuous chemotherapy (Arm A, n=815), continuous chemotherapy with cetuximab (Arm B, n=815), or intermittent chemotherapy (Arm C, n=815) (Adams et al., 2011, Maughan et al., 2011).

COIN-B patients were randomised 1:1 to receive intermittent chemotherapy and cetuximab (Arm D, n=112) or intermittent chemotherapy and continuous cetuximab (Arm E, n=114). For the first 12 weeks, treatments were identical in all patients except from the choice of fluoropyrimidine (n=1,068, 40% received intravenous 5-FU, folinic acid (leucovorin) and oxaliplatin (FOLFOX) and n=1,603, 60% received orally administered capecitabine and intravenous oxaliplatin (XELOX)) together with the randomisation of ± cetuximab (n=1,041, 39% received cetuximab). Patients were aged between 18 and 87 (mean age: 62 years) and 64% were male. Blood DNA samples were prepared from 2,244 patients, all of whom gave fully informed consent for bowel cancer research (approved by REC [04/MRE06/60]). A full breakdown by clinicopathological data is shown in Table 2.2.

2.3.1 Somatic tumour DNA analyses

As described in Maughan et al. (2011) and Smith et al. (2013a), somatic *KRAS*, *BRAF*, *NRAS* and MSI status was determined from formalin-fixed, paraffin embedded (FFPE) tumour samples collected with patient consent. Tumour samples were available for 1,976 patients from COIN and 208 from COIN-B (Smith et al., 2013a, Wasan et al., 2014). Somatic mutations in codons 12, 13, and 61 of *KRAS* and codon 600 of *BRAF* were screened for using pyrosequencing and sequenom. Mutations in codon 12 and 16 of *NRAS* using only sequenom. MSI status was detected using markers BAT-25 and BAT-26 (Smith et al., 2013a).

2.3.2 Germline DNA analyses

DNA samples extracted from the blood of 2,244 patients were genotyped using Affymetrix Axiom Arrays according to the manufacturer's recommendations

Table 2.2. Clinicopathological data of patients by trial arm

		COIN			COIN-B	
Trial & Arm		**A**	**B**	**C**	**D**	**E**
Patients	Total	815	815	815	112	114
	Genotyped and passed QC	579 (71)	616 (76)	583 (72)	85 (76)	85 (75)
Mean Age		62.3	62.9	63.2	61.8	61.9
Sex	Male	390 (67)	410 (67)	376 (64)	48 (56)	46 (54)
	Female	189 (33)	206 (33)	207 (36)	37 (44)	39 (46)
Primary tumour location	Colon	390 (67)	408 (66)	405 (70)	52 (61)	69 (81)
	Rectum	187 (32)	208 (34)	177 (30)	33 (39)	16 (19)
	n/k	2 (<1)	0 (0)	1 (<1)	0 (0)	0 (0)
Number of metastatic sites	0–1	197 (34)	239 (39)	208 (36)	30 (35)	32 (38)
	≥2	382 (66)	377 (61)	375 (64)	55 (65)	53 (62)
Liver-only metastases	Yes	432 (75)	462 (75)	440 (75)	0 (0)	0 (0)
	No	147 (25)	154 (25)	143 (25)	0 (0)	0 (0)
	n/k	0 (0)	0 (0)	0 (0)	85 (100)	85 (100)
Synchronous metastases	Yes	393 (68)	426 (69)	411 (70)	61 (72)	67 (79)
	No	180 (31)	187 (30)	167 (29)	23 (27)	18 (21)
	n/k	6 (1)	3 (<1)	5 (1)	1 (1)	0 (0)
Chemotherapy received	FOLFOX	200 (35)	212 (34)	212 (36)	85 (100)	85 (100)
	XELOX	379 (65)	404 (66)	371 (64)	0 (0)	0 (0)
White blood cell count	<10000 (per L)	404 (70)	442 (72)	399 (68)	73 (86)	63 (74)
	≥10000 (per L)	175 (30)	174 (28)	183 (31)	12 (14)	21 (25)
	n/k	0 (0)	0 (0)	1 (<1)	0 (0)	1 (1)
WHO performance status	0–1	537 (93)	575 (93)	535 (92)	80 (94)	76 (89)
	2	42 (7)	41(7)	48 (8)	5 (6)	9 (11)
Cetuximab administered	Yes	0 (0)	616 (100)	0 (0)	85 (100)	85 (100)
	No	579 (100)	0 (0)	583 (100)	0 (0)	0 (0)
RAS status	mutant	214 (37)	241 (39)	196 (34)	25 (30)	20 (23)
	wild type	261 (45)	268 (44)	293 (50)	52 (61)	60 (71)
	no data	104 (19)	107 (17)	94 (16)	8 (9)	5 (6)
BRAF status	mutant	44 (8)	29 (5)	52 (9)	6 (7)	12 (14)
	wild type	426 (74)	480 (78)	435 (75)	46 (54)	51 (60)
	no data	109 (19)	107 (17)	96 (16)	33 (39)	22 (26)

Response at 12 weeks	Yes	277 (48)	300 (49)	289 (46)	49 (58)	39 (46)
	No	218 (38)	223 (36)	210 (36)	21 (25)	23 (27)
	no data	84 (14)	93 (15)	84 (14)	15 (17)	23 (27)
Median OS (days)		503	496	461	509	527

COIN patients were randomised to receive continuous oxaliplatin and fluoropyrimidine chemotherapy (Arm A), continuous chemotherapy with cetuximab (Arm B), or intermittent chemotherapy (Arm C). COIN-B patients were randomised to receive intermittent chemotherapy and cetuximab (Arm D) or intermittent chemotherapy and continuous cetuximab (Arm E). Data shown for patients that were genotyped and successfully passed quality control. Patients that responded to treatment had either complete or partial response as defined by REICIST 1.0 guidelines. Patients that did not respond had stable or progressive disease. Age: Age at randomisation. OS=Overall survival, n/k=not known, QC=Quality control, FOLFOX=intravenous 5-FU, folinic acid (leucovorin) and oxaliplatin, XELOX=orally administered capecitabine and intravenous oxaliplatin. Percentages are shown in parentheses.

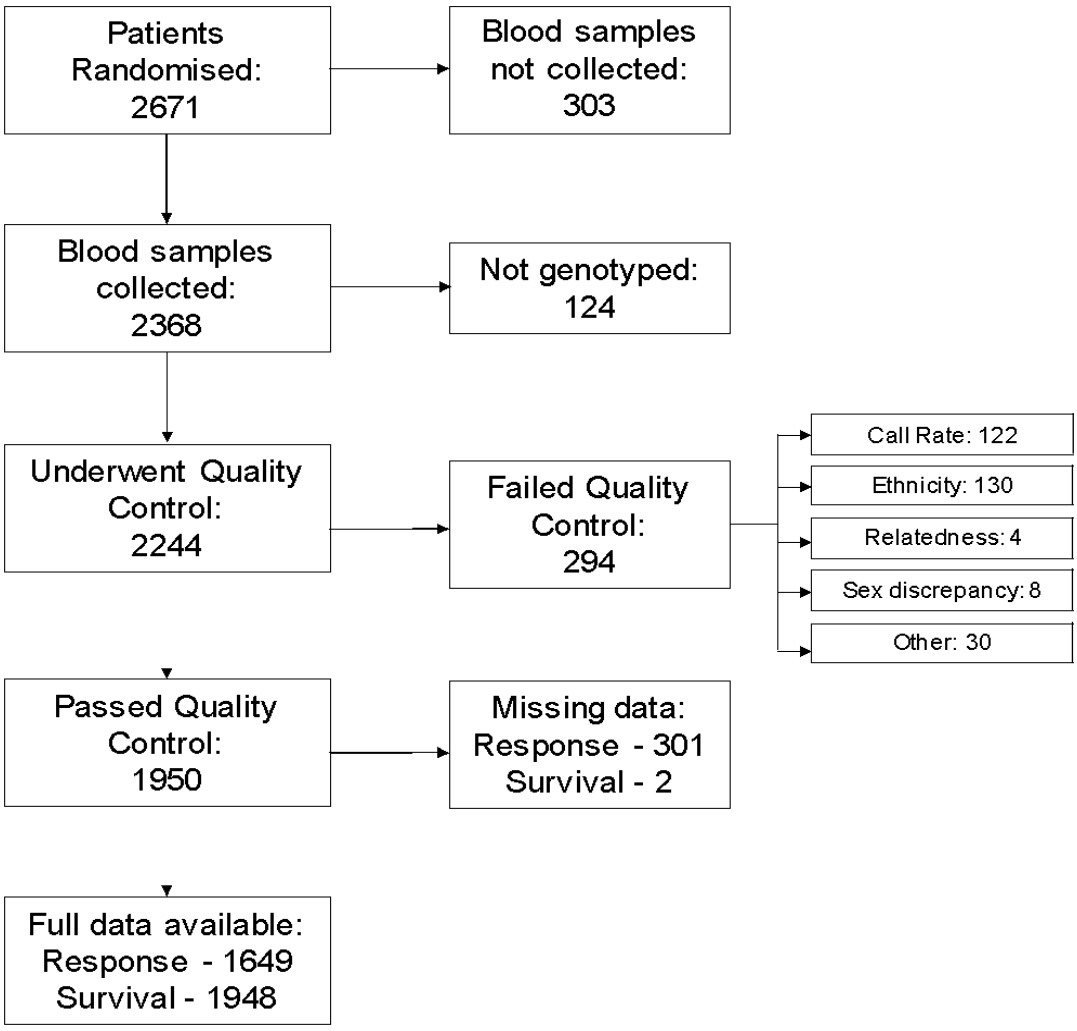

Figure 2.2. CONSORT diagram of COIN/COIN-B patients included in my analyses
Of 2,671 patients enrolled in the COIN and COIN-B trials, 2,244 were genotyped on Axiom arrays, 1,950 passed QC, 1,948 had complete survival data and 1,649 had response data.

(Affymetrix, Santa Clara, CA 95051, USA) at the King Faisal Specialist Hospital and Research Center, Saudi Arabia (under IRB approval 2110033) (Al-Tassan et al. 2015).

Genotyping quality control was tested using duplicate DNA samples with >99% concordance. Individuals were excluded from analysis if they failed one or more of the following thresholds: overall successfully genotyped SNPs <95% (n=122), discordant sex information (n=8), classed as out of bounds by Affymetrix (n=30), duplication or cryptic relatedness (n=4), and non-white European ancestry by Principle Component (PCA)-based analysis (n=130) (Figure 2.2). After quality control (QC), we had whole genome SNP genotyping data on 1,950 patients.

2.3.2.1 Germline DNA quality control

As part of this thesis, further QC was undertaken prior to the analysis of data. Using the *-keep* commands in PLINK version 1.9, SNPs with an infoscore <0.8 (calculated using SNPTEST), genotyping rate <98%, minor allele frequency (MAF) <0.05 and deviated from HWE of $1x10^{-6}$ were excluded from further analyses (Anderson et al., 2010, Purcell et al., 2007). After QC, ~2.8 million SNPs were carried forward for analysis.

2.3.3 Clinical endpoints assessed

Assessment of response was performed at 12 weeks. At this point, patients from all trial arms received identical levels of chemotherapy with or without cetuximab. This time point was also prior to any interruption to treatment for the intermittent therapy arms (C, D and E). Using the RECIST 1.0 guidelines, response was defined as complete or partial response and no response was defined as stable or progressive disease (Therasse et al., 2000). Of the 2,671 patients with complete clinical data, 2,643 had data for their 'best response' with no indication of the time this assessment was performed. However, of these 2,643 with best response data, 476 have no record of assessment at 12 weeks. Therefore the 2,168 patients that had data for response at 12 weeks and could be considered for this investigation. Of these 2,168 patients, 1,649 had complete SNP genotype data and data for response at 12 weeks.

Chi-squared tests were performed to compare patients who were given FOLFOX with patients given XELOX and comparing patients given cetuximab with those who were not. There was no difference in response rates in either chemotherapy type or whether patients received cetuximab (Table 2.3). Due to this homogeneity, all patients with complete response data at 12 weeks across all arms were combined for analyses of response when required unless otherwise stated.

One-way analysis of variants (ANOVA) performed between COIN and COIN-B (Arms A, B and C versus Arms D and E) and between each arm across both cohorts (Arms A, B, C, D and E). There was no difference in OS between both COIN and COIN-B (P=0.25) and across all arms (P=0.49). Due to this lack of heterogeneity and consistency of treatment up until 12 weeks – the endpoint for response, COIN and COIN-B datasets were combined for all analyses.

A Cox regression and log-rank test was performed to determine the relationship between response at 12 weeks and OS in COIN and COIN-B using *survminer* and *ggplot2* (n=1649 patients data for both OS and response at 12 weeks) (Wickham, 2016, Kassambara et al., 2021). There was a significant association between response and better OS (Cox regression HR=0.51, 95% CI=0.45–0.57, $P<2.0x10^{-16}$, Figure 3).

2.4 Statistical analysis

All statistical analyses were two sided and performed using R-version 3.5.2 (CRAN Corporation) – unless otherwise stated (R Core Team, 2018).

2.4.1 Survival analyses

Survival outcomes were analysed in R-version 3.2.5 (R Core Team, 2018). OS was analysed by univariate and multivariate Cox proportional hazards models. Survival curves were plotted using the Kaplan-Meier method and analysed by the log-rank test using the *survival, survminer* and *ggplot2* packages in R (Kassambara et al., 2021, Wickham, 2016).

Table 2.3. Test for heterogeneity between treatment groups and response at 12 weeks for all patients in the COIN/COIN-B dataset

	No Response	Response	All Patients	*X squared*	*P*
Chemotherapy Type					
XELOX	433 (43%)	563 (57%)	996 (100%)		
FOLFOX	262 (40%)	391 (60%)	653 (100%)	1.82	0.18
Cetuximab					
No	428 (43%)	566 (57%)	994 (100%)		
Yes	267 (40%)	388 (60%)	655 (100%)	0.85	0.36

Patients were excluded if they were missing germline genetic data or response data. At the time of data collections, all patients had undergone 12 weeks of continuous oxaliplatin based chemotherapy. Responsive patients were defined as those who had complete or partial response. Non-response was determined as stable or progressive disease as defined by REICIST 1.0 guidelines. *P*-values were calculated through chi-squared tests with 1 degree of freedom.

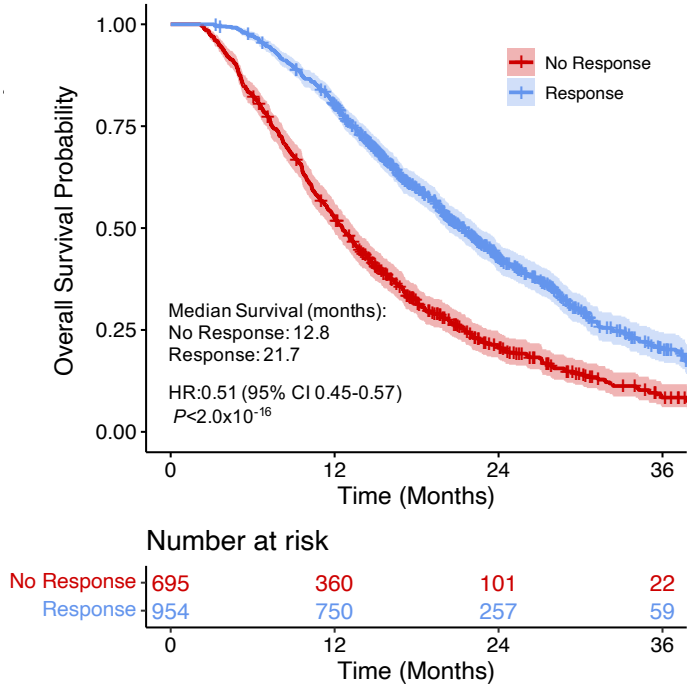

Figure 2.3. Kaplan-Meier demonstrating the difference in overall survival depending on response status (n=1,649)

P calculated using the log-rank test. Patients that responded to treatment had either complete or partial response as defined by REICIST 1.0 guidelines. Patients that did not respond had stable or progressive disease.

2.4.2 Exploratory analyses of clinicopathological factors

Logistic regression analyses for response and selected clinicopathological factors (Chapter 5, Section 2.2.21) was performed in STATA/SE version 16.1. An interaction effect between each factor and cetuximab was calculated by performing an interactive logistic regression. A Forest plot for associations was visualised using the IPDMEDAN package in STATA (Fisher, 2014).

2.4.3 GWAS

Binary variables (response at 12 weeks – response vs no response) were analysed using a logistic regression using the *-logistic* functions in PLINK version 2.0 (Chang et al., 2015). Univariate analysis consisted of only the SNP genotype and the continuous outcome.

A multivariate GWAS for response was performed using the *-logistic* functions in PLINK version 2.0 and adjusting for specific covariates (Chapter 4.2.3.1 and 5).

Analyses were performed on directly genotyped SNPs or those imputed with infoscores of ≥0.8, MAF ≥5% and a HWE of $\geq 1.0 \times 10^{-6}$. Genome wide significance was $P < 5 \times 10^{-8}$, with suggestive association set at $P < 1 \times 10^{-5}$. Visualisation of results was plotted using the R package *qqman* (Turner, 2014).

2.4.4 Power considerations

Power to detect response effect sizes was calculated using the *genpwr* package in R, based upon a predefined power, the *de facto* GWAS significance of $P = 5.0 \times 10^{-8}$ and SNPs with predefined MAFs (5-20%) (Moore and Jacobson, 2020). Power calculations for Cox regressions was performed using the *stpower* function in STATA SE 16.1 or *survSNP* package in R (Owzar et al., 2012).

2.4.5 Gene-based and gene-set analyses

Gene-based and gene-set analyses were performed using MAGMA version 1.8, NCBI 37.3 gene definitions and predefined gene sets (~8000) (de Leeuw et al., 2015),. Gene analysis is based upon multiple linear principal components regression, the *P*-value for this test is derived from the *F*-score of this analysis (Massy, 1965). A Bonferroni correction was used for 20,000 genes to generate an adjusted $P=2.5x10^{-6}$ (Kiezun et al., 2012).

Gene-set analysis was performed using an intercept-only linear regression model, where each gene and *P*-value from the gene analysis are combined to create a normally distributed input for this analysis. A competitive model was used for gene-set analyses, which corrects for gene size for more stringency (de Leeuw et al., 2015). *P*-values from this analysis were corrected to *Q*-scores to take into account false discover rate using the Bioconductor R package *Qvalue* (Storey et al., 2020). Significance was defined as $Q<0.05$.

2.5 Other bioinformatic analyses

Regional association analyses of SNPs was performed using LocusZoom (Pruim et al., 2010). Linkage disequilibrium (LD) of nearby variants was calculated in relation to the sentinel SNP.

Identification of expression quantitative trait loci (eQTLs) was performed using the GTEx project database. This includes data from 838 donors, aged 20–79 years old and 67.1% male. Of these, 84.6% were white, 12.9% African-American, 1.3% Asian, 0.2% American Indian with the remaining donors having unknown heritage. Full details of the methodology for sequencing of project tissue samples can be found at https://gtexportal.org/home/documentationPage. Significance for tissue association was set at $P<1.0x10^{-3}$ (Bonferroni correction for 49 tissue types tested).

2.6 Study design

All analyses performed are retrospective with sample size predetermined by recruitment of patients to the COIN and COIN-B trials. Due to all patients having advanced (stage IV) CRC, no stratification for disease stage was performed. *P*-values in this thesis are uncorrected unless otherwise stated. Analyses in this thesis were performed and reported in accordance with the REMARK guidelines, with all endpoints and sample sizes predefined before analysis (McShane et al., 2005).

3. Chapter 3: Pattern recognition receptor polymorphisms as predictors of oxaliplatin benefit

3.1 Introduction

3.1.1 Pattern recognition proteins in colorectal cancer

In epithelial cancers, including CRC, it has been shown that variation in genes of the innate immune system may reduce the efficacy of cytotoxic cancer treatments (Table 3.1). These treatments may induce immunogenic cell death through PRPs such as Formyl Peptide Receptors and Toll-Like Receptors (Chapter 1, Section 2) (Kroemer et al., 2015).

Previous studies have demonstrated that SNPs in the PRPs *Formyl Peptide Receptor 1* (*FPR1*), *Toll-like Receptor 3* (*TLR3*) and *Toll-like Receptor 4* (*TLR4*) reduce the frequency of presentation of endogenous ligands in dying cells (Tesniere et al., 2010, Vacchelli et al., 2015, Ghiringhelli et al., 2009, Apetoh et al., 2007, Vacchelli et al., 2016). It has been suggested that this lowered ligand presentation reduces the induction of immunogenic cell death which is required for the activity of anthracycline, and most importantly for CRC patients, oxaliplatin (Ghiringhelli et al., 2009, Tesniere et al., 2010, Vacchelli et al., 2015).

Variations in these PRPs have been analysed in trials exploring the effect of SNPs on survival outcome (Table 3.1).

Following identification of resistance to chemotherapy (likely facilitated by a reduced antitumour immunity) in tumor-bearing Fpr1 (-/-) mice, Vacchelli et al. (2015) showed a 'variation in *FPR1*' (rs867228) was associated with reduced OS (HR=0.63, $P<0.01$) and PFS (HR=0.56, $P<0.05$) in CRC patients. This variation was shown to reduce both OS and metastatic-free survival (MFS) in breast cancer patients undergoing treatment by anthracycline (Vacchelli et al., 2015). rs867228 (Chromosome 19:51745958 (GRCh38), *FPR1* c.1037A>C) has been described as a LOF variant due to the resulting amino acid substitution (Glu346Ala) in the intracellular C-terminus of some

isoforms of FPR (Seifert and Wenzel-Seifert, 2003, Sherry et al., 2001). Site directed mutagenesis studies such as Wenzel-Seifert and Seifert (2003) show this substitution influences the dimerization of *FPR* isoform 26 resulting in reduced constitutive activity.

A SNP in *TLR4* (rs4986790) alone has been demonstrated to reduce OS (HR=0.72, *P*=0.05) and PFS (HR=0.73, *P*<0.05) in CRC patients (Tesniere et al., 2010). rs4986790 (Chromosome 9:117713024 (GRCh38), TLR4 c.896A>G) causes an amino acid substitution (Asp299Gly) (Sherry et al., 2001). Functional assays (including overexpression studies) have shown that this substitution disrupts the extracellular domain of *TLR4* which is likely to disturb its transportation to the cell membrane (Arbour et al., 2000, Schröder and Schumann, 2005). Cell lines transfected with lentiviral constructs of *TLR4* with this amino acid substitution were found to have a significantly lower response to stimuli compared to wild-type *TLR4* and *TLR4* Thr399Ile, indicating the specific role of Asp299Gly in the activity of *TLR4* (Long et al., 2014).

Vacchelli et al. (2015) also indicated that the 'mutational background' may play a role in the relationship between rs867228 and survival outcomes. It has been reported that this variation only has an effect on survival (rs867228 A>C is significantly associated with DFS) in patients with functional *TLR3* or *TLR4* (or both) (Vacchelli et al., 2015). This is likely due to the fact that these genes are found within the same pathway (Chapter 1, Section 2 and Figure 1.8).

3.1.2 Validation of previous studies

Oxaliplatin is the most common form of treatment for CRC and the previously reported polymorphisms (Table 3.1) are common (>5% allelic frequency) in individuals of European descent. If these associations can be confirmed, this could be important for current clinical practice. Validated variations in *FPR1, TLR3 and TLR4* could be used as prognostic biomarkers facilitating the targeted use of oxaliplatin which could result in less harm due to the toxicity of these therapies and savings for health-care providers (Hoff et al., 2012).

Table 3.1. Previously published data for Pattern Recognition Proteins and Cancer Clinical Outcomes

Study	Variant	Cancer type	N	Primary Endpoint	Inheritance Model	HR (95% CI)	*P*
Vacchelli et al. (2015)	rs867228 *FPR1* c.1037A>C	Early-stage breast	731	OS	Dominant	1.4 (1.1–2.1)	<0.01
			731	MFS	Dominant	1.6 (1.2–2.1)	<0.01
		mCRC	311	OS	Recessive	2.1 (1.3–3.6)	<0.01
			311	PFS	Recessive	1.9 (1.1–3.1)	<0.03
Chen et al. (2015)	rs3775291 *TLR3* c.1234C>T	Breast	715	RFS	Recessive	3.5 (2.0–6.3)	<0.01
Castro et al. (2011)	rs3775291 *TLR3* c.1234C>T	CRC	565	CRC-Specific Survival	Recessive	2.00 (1.1–3.3)	0.03
Tesniere et al. (2010)	rs4986790 *TLR4* c.896A>G	mCRC	338	OS	Multivariate	1.4 (1.0–1.9)	0.05
			338	PFS	Multivariate	1.4 (1.0–2.0)	<0.05
Apetoh et al. (2007)	rs4986790 *TLR4* c.896A>G	Node positive breast	280	MFS	Dominant	Not reported	0.03

OS=Overall Survival, MFS=Metastasis-free Survival, PFS=Progression-free Survival, RFS=Recurrence-free Survival

3.1.3 Hypothesis and aims

The main hypothesis of this chapter is that previously reported associations between PRPs and survival and response outcomes will be observed in aCRC patients from COIN and CON-B with confirmation of findings from cases from an earlier phase trial, Short Course in Oncology Therapy (SCOT) (Tesniere et al., 2010, Vacchelli et al., 2015, Iveson et al., 2018). The specific aims are:

- Validate previously reported associations of treatment outcomes for oxaliplatin and SNPs in *FPR1, TLR3* and *TLR4*.
    - Perform survival and response regression analyses on all patients treated with oxaliplatin.
    - Perform survival regression analyses on patients with specific PRP variation backgrounds.
- Investigate variations in other PRPs for associations with oxaliplatin treatment outcomes.

3.2 Methods

3.2.1 DNA extraction and Quality control

Full details for DNA extraction and QC for COIN and COIN-B genotyping can be found in methods chapter (Chapter 2, section 2). Of the 2,671 patients recruited, 2,365 patients underwent genotyping of which 1,950 patients passed QC.

3.2.2 Statistical Analyses

The two phenotypes used for the statistical analysis of COIN and COIN-B were OS and radiological response at 12 weeks. OS is defined as the number of days from randomisation to death from any cause or time of last contact with the clinic (n=1,948, two patients with incomplete data excluded). Response at 12 weeks is a binary outcome – response (complete or partial response as defined by RECIST 1.0 guidelines) or non-response (stable or progressive disease) and was chosen due to the consistency of treatment across all arms up until this time point (n=1649, 301 patients excluded due to missing data) (Therasse et al., 2000).

3.2.2.1 *FPR1*, *TLR3* and *TLR4*

rs867228 (*FPR1*), rs3775291 (*TLR3*) and rs4986790 (*TLR4*) were analysed individually in R-version 3.2.5 (R Core Team, 2018). OS was analysed by univariate and multivariate (SNP genotype, baseline demographics, treatment type and selected prognostic factors associated with CRC) Cox proportional hazards models, under three different genetic models – dominant, recessive and additive using the survival package in R-version 3.2.5 (Therneau, 2020).

HRs detectable in multivariate analyses (Table 3.2) are based on a two-sided significance (α) of 0.05 and a power (1-β) of 80%. These were calculated using STATA version 13 (StataCorp, College Station, Texas, USA).

Table 3.2. Detectable effect sizes with 80% power in multivariate Cox regression analyses with overall survival as a primary endpoint

	Cases	Events	HR Previously Reported	Detectable HR
rs867228 (*FPR1* c.1037C)	1,336	970	2.12	
Recessive				1.22
Dominant				1.37
rs3775291 (*TLR3* c.1234T)	1,563	1,150	N/A	
Recessive				1.35
Dominant				1.18
rs4986790 (*TLR4* c.896G)	1,563	1,150	1.36	
Recessive				3.23
Dominant				1.35

Detectable hazard ratios (HRs) for each SNP and inheritance model take into account SNP prevalence, sample size and probability of failure. All calculations are based on a power (1-β) of 0.8 and a two-sided significance (α) of 0.05. Reported HRs represent previously reported data regarding these SNPs and overall survival as a primary endpoint. HR for rs3775291 not previously reported. N/A=not applicable, HR not previously reported.

Covariables for multivariate analysis were pre-specified based on data availability and the proportion of missing values with no further selection (e.g., backwards elimination) being performed. The genotypes were manually recoded to reflect these inheritance models for example – rs867228 (A/C with C as the 'LOF allele') the dominant model is AA [0] vs CC and CA [1], recessive AA and CA [0] vs CC [1] and the additive model implies AA [0] vs CA [1] vs CC [2].

Survival curves for SNP genotypes were plotted using the Kaplan-Meier method and analysed by the log-rank test using the survival and ggplot2 in R-version 3.2.5 (Therneau, 2020, Wickham, 2016). *P*-values for individual predictors in Cox models were calculated by the Wald test.

As response at 12 weeks was a binary outcome, either Chi-squared tests or a Fisher's exact test, in the case of rs4986790 where the assumptions for Chi-squared could not be met, were used to perform a univariate analysis for the association of each SNP using the base functions in R-version 3.2.5 (R Core Team, 2018).

3.2.3 Confirmatory Study

Genotyping and survival analysis (OS and DFS) on the pattern recognition SNPs rs867228, rs3775291 and rs4986790 were also investigated in an additional clinical trial, SCOT (Iveson et al., 2018). SCOT's primary aim was to determine if 12 weeks of oxaliplatin-based chemotherapy was more efficient than the standard 24 weeks of treatment for high risk stage II, or stage III CRC. 2,939 patients had their DNA extracted using standard methods, passed genomic QC and were phased using SHAPEIT and imputed using IMPUTE2 (Delaneau et al., 2011, Marchini et al., 2007). rs867228 was imputed with an info score of 0.95, with rs3775291 and rs4986790 being directly genotyped. Due to lack of data, no response analysis was performed.

Methodologies for survival analyses were the same as described in section 2.2.1, however, due to differences in trial data collection, different covariates are used for the multivariate analysis. These included nodal status (N0 vs N1 vs N2) and treatment duration (24 vs 12 weeks) which differed for COIN and COIN-B.

3.2.4 Follow up analyses

Follow-up analyses of an extended profile of genes were performed on 12 genes from four families related to the PRPs previously explored. SNPs to be studied were sourced from SNPedia (https://www.snpedia.com) and dbSNP (https://www.ncbi.nlm.nih.gov/snp/) and had a MAF>5% as calculated in PLINK (Cariaso and Lennon, 2012, Sherry et al., 2001, Chang et al., 2015). OS was analysed using GenABEL in R-version 3.5.2 (CRAN corporation) by Cox regression in an additive model (Karssen et al., 2016, R Core Team, 2018). Response at 12 weeks was analysed using a logistic regression in an additive model in PLINK version 1.9 (Chang et al., 2015).

P-values for these tests were adjusted for multiple testing using a Bonferroni correction to a new significance threshold of $P=8.6 \times 10^{-5}$ (Bland and Altman, 1995).

3.3 Results

3.3.1 Cohort background

Of the 2,445 patients in COIN/COIN-B who consented to bowel cancer research, 1,948 and 1,649 patients had complete SNP genotype and survival and response data respectively. For the 1,948 patients with survival data, median follow up time was 23.2 months. At this time, 1,453 deaths had occurred.

3.3.2 Statistical Power

HRs detectable for multivariate analyses are shown in Table 3.2. In comparison to previous studies, the power to detect an identical effect size for rs867228 and OS in a recessive model was 100%. The power to detect the previously reported effect size in the dominant model for rs4986790 was 96%.

3.3.3 Genotype distributions and alleles frequencies

rs867228 was imputed in 1,672 patients. rs3775291, rs4986790 and rs4986791 were directly genotyped in 1,948 patients. Distribution of these SNP genotypes is shown in Table 3.3. Allelic frequencies for all four SNPs were found to be concordant to the general population as described in UK10K, ExAC and EVS (Walter et al., 2015, Lek et al., 2016, Tennessen et al., 2012). rs4986790 and rs4986791 were found to have similar allelic frequencies (Table 3.3) and are in LD in European populations (D'=0.9814, $R^{2=}$0.9455) (Machiela and Chanock, 2015). Therefore, of these two SNPs, only rs4986790 was carried forwards for further investigation. All SNPs were found to be in HWE (P<1x10^{-4}).

Table 3.3. Distribution of genotypes for variants of *FPR1*, *TLR3* and *TLR4* in the COIN/COIN-B cohort

Variant	N	%
All Samples	1,948	100
FPR1 rs867228 genotype		
AA	49	2.5
AC	444	22.8
CC	1,179	60.5
Unknown	276	14.2
TLR3 rs3775291 genotype		
CC	934	47.9
CT	810	41.6
TT	204	10.5
Unknown	0	0.0
TLR4 rs4986790 genotype		
AA	1,744	89.5
AG	200	10.3
GG	4	0.2
Unknown	0	0.0
TLR4 rs4986791 genotype		
CC	1,726	88.6
CT	218	11.2
TT	4	0.2
Unknown	0	0.0

Percentages (%) are of all successfully genotyped samples with survival data (n=1948). Genotypes for *FPR1* (rs867228) were derived via imputation score with ambiguous genotypes being attributed as unknown. All other SNPs were directly genotyped.

3.3.4 Known prognostic factors in COIN and COIN-B

Association tests (Cox regression) for known prognostic factors were performed (Table 3.4) to ensure that multivariate analyses included key prognostic factors. Disease site (HR=0.86, 95% CI=0.76–0.98, *P*=0.02), WHO performance status (HR=1.50, 95% CI=1.20–1.87, $P=3.6\times10^{-3}$), primary tumour resection status (HR=0.74, 95% CI=0.65–0.83, $P=1.6\times10^{-7}$), white cell count (HR=1.64, 95% CI=1.43–1.87, $P=4.8\times10^{-13}$) and tumour biomarker mutation status (*KRAS* mutation HR=1.51, 95% CI=1.33–1.72, $P=1.9\times10^{-10}$; *NRAS* mutation HR=1.55, 95% CI=1.17–2.06, $P=2.3\times10^{-3}$; *BRAF* mutation HR=2.43, 95% CI=1.98–2.97; $P<2.0\times10^{-16}$) were found to be significantly associated with OS (*P*<0.05).

3.3.5 Pattern recognition SNPs and overall survival

When performing a univariate analysis, there was no association between rs867228, rs3775291 or rs4986790 and OS regardless of inheritance model (Table 3.5). Similarly, under multivariable cox regression analyses, none of these SNPs influenced OS (Table 3.6). This is also seen in the results of Kaplan-Meier analyses and two-sided log rank tests for each SNP individually (Figure 3.1A-C, rs867228 *P*=0.78, rs3775291 *P*=0.31 and rs4986790 *P*=0.60).

3.3.6 Association of *FPR1* on different pattern recognition backgrounds and clinical outcome

Further analyses were performed to examine the relationship between rs8867228 and OS after stratification for *TLR3* (rs3775291) and *TLR4* (rs4986790) variation (Figure 3.2A-F).

There was no association between rs8867228 and OS when analysed on *TLR3* (*P*=0.91, log rank test) or *TLR4* (*P*=0.48) wild-type backgrounds nor on *TLR3* (P=0.53) and *TLR4* (P=0.40) mutant backgrounds.

Table 3.4. Prognostic associations of factors included in the multivariate analysis of *FPR1*, *TLR3* and *TLR4* SNPs in the combined COIN/COIN-B cohort (1,563 cases, 1,150 deaths)

	HR (95% CI)	*P*
Age	1.01 (1.00–1.01)	0.10
Sex		
Male	1.0	
Female	1.08 (0.95–1.22)	0.23
Disease site		
Colon	1.0	
Rectum	0.86 (0.76–0.98)	0.02
WHO performance status		
0–1	1.0	
2	1.50 (1.20–1.87)	3.6×10^{-4}
Primary tumour resected		
No	1.0	
Yes	0.74 (0.65–0.83)	1.6×10^{-7}
Local Recurrence	0.88 (0.67–1.16)	0.36
White cell count		
<10,000/mcL	1.0	
≥10,000/mcL	1.64 (1.43–1.87)	4.8×10^{-13}
KRAS mutation status		
Wild-type	1.0	
Mutant	1.51 (1.33–1.72)	1.9×10^{-10}
NRAS mutation status		
Wild-type	1.0	
Mutant	1.55 (1.17–2.06)	2.3×10^{-3}
BRAF mutation status		
Wild-type	1.0	
Mutant	2.43 (1.98–2.97)	$<2.0 \times 10^{-16}$
Cetuximab treatment		
No	1.0	
Yes	1.05 (0.92–1.20)	0.45
Chemotherapy regimen		
CAPOX	1.0	
FOLFOX	0.95 (0.84–1.07)	0.41
Chemotherapy arm		
Continuous	1.0	
Intermittent	1.07 (0.94–1.22)	0.31

Multivariate Cox proportional hazard regression analyses performed with OS as the primary end point. Analysis excludes cases with missing values for any prognostic factor. Hazard ratios for each variable are adjusted for all other covariates listed. Hazard ratios for reference groups are listed as 1.0.

Table 3.5. Univariate analyses of overall survival in COIN/COIN-B by SNP

	N	OS events	HR (95% CI)	*P*
rs867228 (*FPR1* c.1037C)	1,672	1,241		
Additive			1.03 (0.93–1.14)	0.60
Recessive			0.98 (0.71–1.37)	0.93
Dominant			1.04 (0.92–1.18)	0.52
rs3775291 (*TLR3* c.1234T)	1,948	1,453		
Additive			0.98 (0.91–1.06)	0.67
Recessive			1.07 (0.90–1.26)	0.45
Dominant			0.94 (0.86–1.05)	0.31
rs4986790 (*TLR4* c.896G)	1,948	1,453		
Additive			1.03 (0.88–1.21)	0.71
Recessive			1.65 (0.61–4.40)	0.31
Dominant			1.02 (0.86–1.20)	0.81

Cox regression analyses were performed with OS as the primary endpoint and individual SNP genotype as the only variable. All analyses were performed on cases with no missing data. Hazard ratios demonstrate the associated risk with loss of the functional allele for each SNP.

Table 3.6. Multivariate analyses of overall survival in COIN and COIN-B by SNP

	N	OS events	HR (95% CI)	*P*
rs867228 (*FPR1* c.1037C)	1,336	970		
Additive			0.99 (0.88–1.12)	0.91
Recessive			0.92 (0.63–1.34)	0.66
Dominant			1.05 (0.90–1.23)	0.53
rs3775291 (*TLR3* c.1234T)	1,563	1,150		
Additive			0.97 (0.89–1.06)	0.56
Recessive			1.08 (0.90–1.31)	0.41
Dominant			0.93 (0.83–1.04)	0.20
rs4986790 (*TLR4* c.896G)	1,563	1,150		
Additive			1.10 (0.91–1.33)	0.31
Recessive			2.91 (0.93–9.12)	0.07
Dominant			1.08 (0.90–1.31)	0.40

Cox regression analyses were performed with OS the primary endpoint. Hazard ratios were calculated using the minor allele for each SNP genotype alongside multiple prognostic variables. Prognostic factors were – age, sex, disease site (colon or rectum), World Health Organisation (WHO) performance status (0 and 1 or 2), resection of the primary tumour (unresected, resected or local recurrence), tumour mutation status (*KRAS*, *NRAS,* and *BRAF* – wild-type or any mutation), white blood cell count (<10,000 cells/mcL or ≥10,000 cells/mcL), presence or absence of the anti-EFGR antibody – cetuximab, chemotherapy regimen (FOLFOX or CAPOX) and schedule (intermittent or continuous). Individuals with missing data for either OS or any prognostic factor or both were excluded from the analysis.

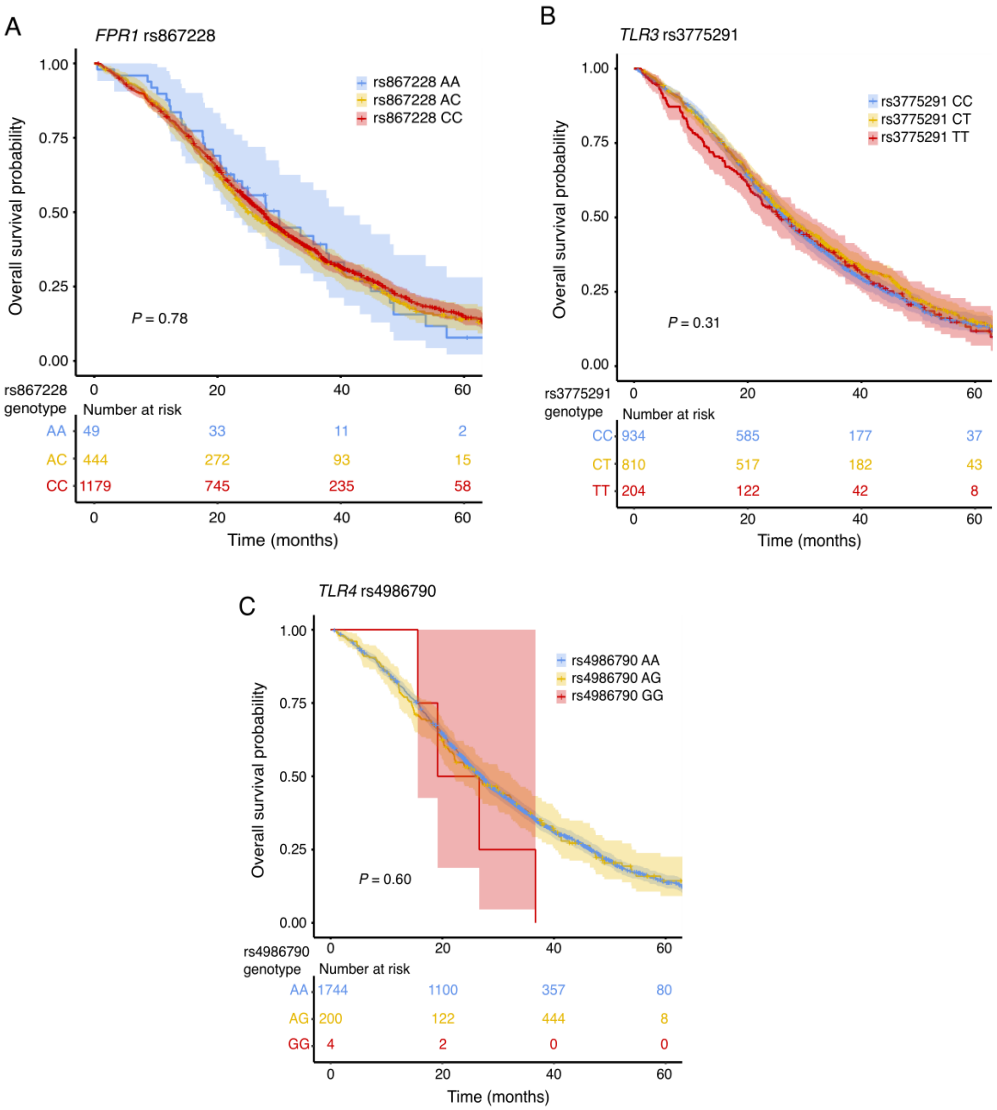

Figure 3.1. Overall survival in relation to loss of function SNPs in *FPR1*, *TLR3* and *TLR4* in COIN/COIN-B

Kaplan-Meier survival curves for OS in combined COIN/COIN-B cohort for 3 different by pattern recognition receptor SNPs. Curves are shown for SNPs (A) in FPR1 (rs867228 A>C) (B) TLR3 (rs3775291 C>T) and (C) TLR4 L (rs4986790 A>G). Due to rs4986791 being in strong LD with rs377529 (C), results were near identical and therefore not shown. Comparison of all groups using the log-rank test produced the *P*-values shown with shaded areas representing 95% CIs for corresponding survival curves. Comparison of all groups using the log-rank test produced the P-values shown. Figure originally from Gray et al. (2019) and reproduced under Creative Commons CC-BY license.

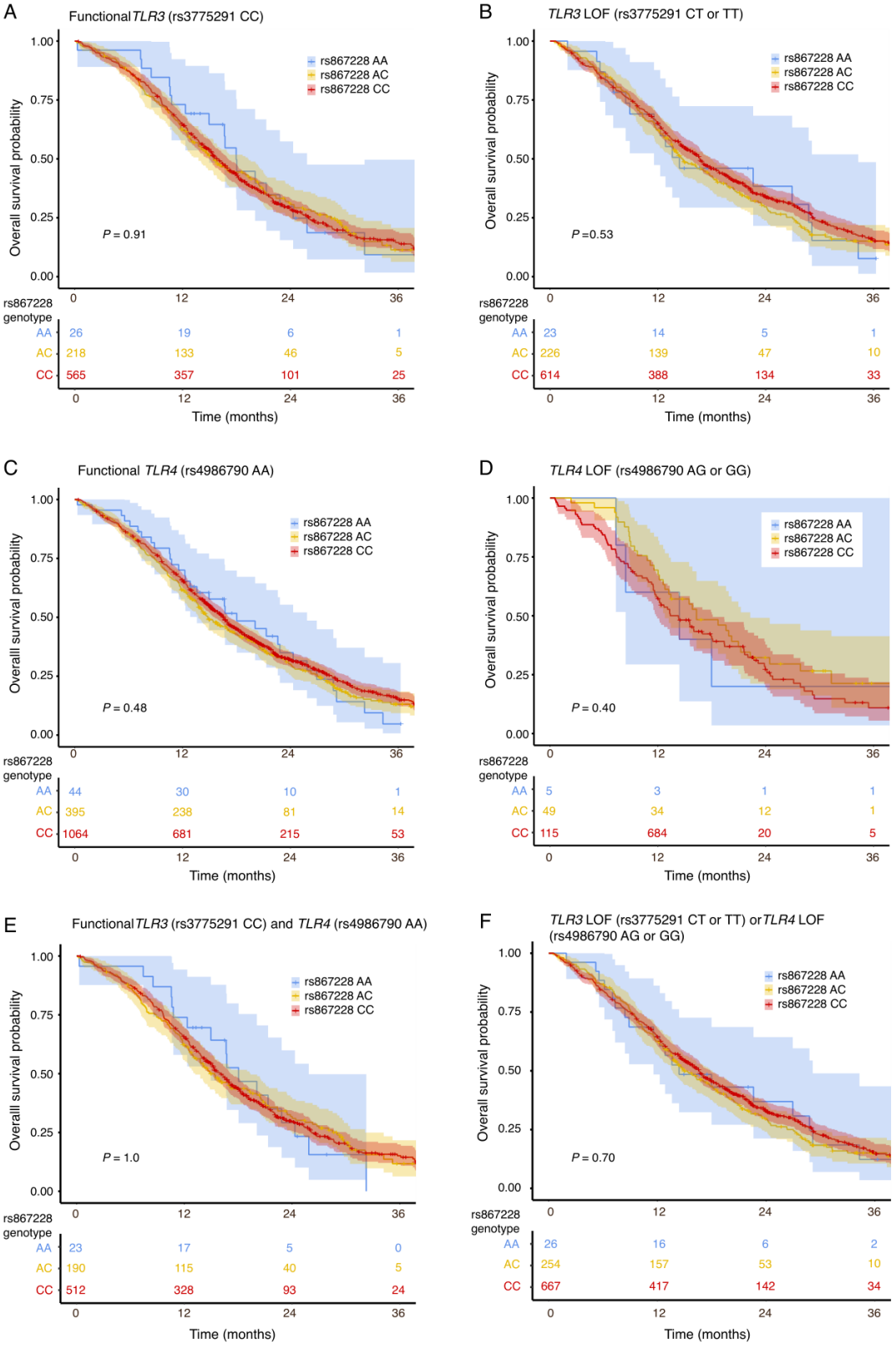

Figure 3.2. Association of *FPR1* (rs867228) with overall survival according to *TLR3* and *TLR4* background in combined COIN/COIN-B cohorts

Kaplan-Meier survival curves showing OS for patients in combined COIN/COIN-B cohorts by *FPR1* (rs867228 A>C), *TLR3* and *TLR4* status. Upper panels show associations in relation to *TLR3* status – either homozygous for the reference allele (rs3775291 C) (A) or heterozygous or homozygous for variation in *TLR3* (SNP rs3775291 C>T) (B). Middle panels show associations of rs867228 and OS in relation to TLR4 status – either homozygous for the rs4986790 wild-type allele (C), or heterozygous and homozygous for the a *TLR4* SNP, rs4986790 (TLR4 c.896A>G) (D). Lower panels show associations of rs867228 genotype in *TLR3* and *TLR4* without variation – with cases of major alleles at both *TLR3* and *TLR4* (E) and with variation at either locus within *TLR3* and *TLR4* (F). Shaded areas represent 95% CIs P-values indicate comparison of all groups by the log-rank test. Figure reproduced from Gray et al. (2019) under Creative Commons CC-BY license.

3.3.7 Pattern recognition SNPs and response to oxaliplatin-based chemotherapy

There was no significant association with response to oxaliplatin at 12 weeks in COIN and COIN-B and rs867228, rs3775291 and rs4986790 genotypes (Table 3.7).

3.3.8 Confirmatory Cohort

2,929 patients were successfully genotyped from the SCOT cohort and independently investigated for associations between rs867228, rs3775291 and rs4986790 and DFS and OS.

SCOT had median follow-up of 36.8 months at which point 186 deaths had occurred. When comparing the SCOT and COIN and COIN-B cohorts, there was a statistically significant difference in some of the baseline demographics including distribution of age (P<0.01, Student's t test) and sex (P<0.01, χ^2 test). As with COIN and COIN-B, SCOT cohort genotyping resulted in allelic frequencies consistent with the general population as described in EVS, ExAC and UK10K (Tennessen et al., 2012, Lek et al., 2016, Walter et al., 2015).

As seen in COIN and COIN-B, rs4986790 and rs4986791 were found in strong LD (D' = 0.99 and r^2 = 0.93) in the SCOT cohort so only rs4986790 used for downstream analysis (Gray et al., 2019).

The power to detect a HR of 2.1 for OS in a recessive model for rs867228 in SCOT was 99%. The power to detect a HR of 1.4 for OS in in a dominant model for rs4986790 was 31%. When performing univariate and multivariate analyses, there was no association for rs867228, rs3775291, or rs4986790 with OS (Table 3.8).

Table 3.7. Univariate analysis of response data for all patients in COIN and COIN-B

	No Response (%)	Response (%)	χ-squared	*P*
rs867228 (*FPR1* c.1037A>C)				
AA	19 (3.1)	22 (2.7)		
AC	166 (27.2)	221 (27.3)		
CC	425 (69.7)	568 (70.0)		
All patients	610 (100.0)	811 (100.0)	0.20	0.90
rs3775291 (*TLR3* c.1234C>T)				
CC	329 (47.3)	417 (49.4)		
CT	293 (42.2)	382 (40.0)		
TT	73 (10.5)	101 (10.6)		
All patients	695 (100.0)	954 (100.0)	0.79	0.68
rs4986790 (*TLR4* c.896A>G)				
AA	628 (90.4)	857 (89.8)		
AG	67 (9.6%)	95 (10.0%)		
GG	0 (0.0%)	2 (0.2%)		
All Patients	695 (100.0%)	954 (100.0%)		0.64

Data was collected after 12 weeks of treatment at which point all patients had undergone 12 weeks of continuous oxaliplatin based chemotherapy. Responsive patients are defined as those who had complete or partial response. Non-response is determined as stable or progressive disease as defined by REICIST 1.0 guidelines. *P*-values are calculated through chi-squared tests or in the case of rs4986790, a Fisher's exact test.

Table 3.8. Analysis of survival data performed on SCOT cohort

	N	OS events	HR (95% CI)	*P*
rs867228 (*FPR1* c.1037C)	2,728	167		
Univariate				
Additive			1.09 (0.83–1.44)	0.53
Recessive			1.07 (0.77–1.49)	0.67
Dominant			1.40 (0.57–3.41)	0.46
Multivariate				
Additive			1.10 (0.84–1.47)	0.48
Recessive			1.10 (0.79–1.53)	0.56
Dominant			1.32 (0.54–3.22)	0.54
rs3775291 (*TLR3* c.1234T)	2,924	186		
Univariate				
Additive			1.15 (0.93–1.34)	0.29
Recessive			1.46 (0.92–2.32)	0.11
Dominant			1.12 (0.84–1.49)	0.44
Multivariate				
Additive			1.13 (0.91–1.41)	0.27
Recessive			1.32 (0.83–2.10)	0.24
Dominant			1.12 (0.84–1.49)	0.44
rs4986790 (*TLR4* c.896G)	2,929	186		
Univariate				
Additive			0.89 (0.57–1.39)	0.62
Recessive			1.93 (0.48–7.76)	0.36
Dominant			0.83 (0.51–1.35)	0.45
Multivariate				
Additive			0.87 (0.56–1.36)	0.54
Recessive			1.82 (0.44–7.40)	0.40
Dominant			0.81 (0.50–1.32)	0.39

Data was produced by D. Church and used Cox regression analyses with OS and DFS as the primary endpoint but only OS is shown. All analyses were performed on all cases with no missing data. HRs demonstrate the associated risk with loss of the functional allele for each SNP. Multivariable analysis was performed with adjustments for age, sex, disease site (colon vs rectum), primary tumour 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).

3.3.9 Extended analysis of pattern recognition protein SNPs

An analysis of variants within FPR, TLR and related families and their relationship with OS and response at 12 weeks was performed. Variants from 12 genes in four families (Table 3.9) were selected based on data availability. Variants with a MAF<5% were excluded due to inadequate power; using only common variants (MAF>5%) results in a power of >80% when detecting a HR of 1.25 with a significance of P<0.05.

In total, 59 SNPs were analysed. Nine variants from five genes (*IFIH1, TLR9, DDX58, NOD2* and *FPR3*) were nominally associated with OS (P<0.05; Table 3.10); however, none were significant after correction for multiple testing (P<8.6 x10^{-5}). Three variants in two genes (*TLR2* and *DDX58*) were nominally associated with response (Table 3.11), but neither remained significant after adjustment for multiple testing (Bland and Altman, 1995).

Table 3.9. Genes and SNPs found in Pattern Recognition Proteins

	Gene	Chromosome	SNPs
Toll Like Receptors			
	TLR1	4	rs4543123, rs4833095, rs5743551, rs5743557, rs5743565, rs5743592, rs5743611
	TLR2	4	rs11938228, rs1816702, rs1898830, rs201786064, rs3804099, rs3804100, rs4696480, rs76112010, rs7656411, rs7696323
	TLR5	1	rs2072493, rs5744174. rs851139
	TLR6	4	rs1039559, rs3775073, rs5743810, rs5743827
	TLR7	X*	
	TLR8	X*	
	TLR9	3	rs352140, rs352139, rs5743836
	TLR10	4	rs4543123, rs4833095, rs5743551, rs5743557, rs5743565, rs5743592, rs5743611
RIG-I-like receptors			
	DDX58	9	rs10813831, rs3205166, rs56309110, rs669260, rs659527, rs10813821, rs3739674, rs11795343, rs1133071
	IFIH1	2	rs1990760, rs3747517
Nucleotide-binding oligomerisation domain-containing proteins			
	NOD1	7	rs2075818, rs2075820
	NOD2	16	rs1077861, rs2066842, rs2066843, rs2067085, rs2076756, rs3135499, rs3135500, rs5743289, rs8057341
Formyl Peptide Receptors			
	FPR2	19	rs11666254, rs4802859
	FPR3	19	rs17695224

Genes on the X chromosome (indicated by *) were not investigated due to lack of availability of genomic data in the COIN/COIN-B dataset. SNPs included are those found in germline dataset with a MAF of >5%.

Table 3.10. Univariate analyses of OS for *TLR*, *FPR* and *NOD* SNPs

SNP	N	HR (95% CI)	*P*	SNP	N	HR (95% CI)	*P*
TLR5				*TLR10*			
rs2072493	1,758	0.94 (0.85–1.05)	0.30	rs4833095	1,649	1.06 (0.89–1.26)	0.53
rs5744174	1,948	1.05 (0.97–1.13)	0.23	rs5743551	1,613	1.06 (0.89–1.24)	0.50
rs851139	1,814	1.05 (0.97–1.14)	0.21	rs5743557	1,625	1.02 (0.85–1.24)	0.81
IFIH1				rs5743565	1,628	1.04 (0.86–1.26)	0.71
rs1990760	1.948	1.01 (0.94–1.09)	0.15	rs5743592	1,640	1.03 (0.85–1.25)	0.76
rs3747517	1,948	1.00 (0.93–1.09)	0.04	rs5743611	1,604	0.94 (0.72–1.22)	0.63
TLR9				*NOD1*			
rs352140	1,811	0.93 (0.86–1.00)	0.06	rs2075818	1,817	1.02 (0.93–1.02)	0.64
rs352139	1,711	1.09 (0.87–1.39)	0.42	rs2075820	1,814	1.02 (0.94–1.12)	0.27
rs5743836	1,783	1.05 (0.95–1.18)	0.33	*DDX58*			
TLR1				rs10813831	1,948	0.92 (0.85–1.00)	0.06
rs4543123	1,903	1.01 (0.93–1.11)	0.72	rs3205166	1,882	0.92 (0.84–1.00)	0.04
rs4833095	1,948	1.00 (0.91–1.10)	0.97	rs56309110	1,937	1.08 (0.99–1.18)	0.08
rs5743551	1,906	1.01 (0.92–1.11)	0.83	rs669260	1,936	1.01 (0.91–1.12)	0.83
rs5743557	1,920	1.00 (0.90–1.11)	0.96	rs659527	1,927	1.01 (0.94–1.09)	0.80
rs5743565	1,922	1.00 (0.90–1.10)	0.96	rs10813821	1,948	0.99 (0.91–1.06)	0.69
rs5743592	1,938	1.00 (0.91–1.11)	0.93	rs3739674	1,928	1.00 (0.93–1.09)	0.90
rs5743611	1,897	1.07 (0.93–1.23)	0.35	rs11795343	1,927	0.95 (0.87–1.02)	0.15
TLR2				rs1133071	1,933	1.05 (0.97–1.14)	0.21
rs11938228	1,948	0.96 (0.89–1.03)	0.27	*NOD2*			
rs1816702	1,948	1.03 (0.91–1.16)	0.60	rs1077861	1,921	0.95 (0.88–1.03)	0.21

rs1898830	1,907	1.01 (0.94–1.10)	0.75	rs2066842	1,923	0.93 (0.86–1.02)	0.10
rs201786064	1,863	1.07 (0.97–1.19)	0.19	rs2066843	1,925	0.94 (0.87–1.02)	0.12
rs3804099	1,948	1.05 (0.97–1.12)	0.23	rs2067085	1,889	1.08 (1.01–1.17)	0.03
rs3804100	1,948	0.98 (0.85–1.13)	0.79	rs2076756	1,948	0.94 (0.87–1.02)	0.19
rs4696480	1,905	0.98 (0.91–1.05)	0.55	rs3135499	1,932	1.09 (1.01–1.17)	0.02
rs76112010	1,932	1.01 (0.94–1.09)	0.75	rs3135500	1,948	1.09 (1.01–1.18)	0.02
rs7656411	1,920	1.05 (0.97–1.14)	0.25	rs5743289	1,948	0.93 (0.85–1.03)	0.19
rs7696323	1,948	1.01 (0.94–1.09)	0.79	rs8057341	1,809	0.99 (0.91–1.07)	0.86
TLR6				*FPR2*			
rs1039559	1,884	0.95 (0.89–1.03)	0.21	rs11666254	1,948	1.00 (0.92–1.08)	0.90
rs3775073	1,855	0.94 (0.87–1.02)	0.15	rs4802859	1,920	1.00 (0.92–1.08)	0.96
rs5743810	1,948	1.05 (0.98–1.13)	0.16	*FPR3*			
rs5743827	1,863	0.96 (0.89–1.04)	0.35	rs17695224	1,948	0.89 (0.82–0.97)	0.01
rs4543123	1,903	1.02 (0.93-1.11)	0.72				

Additive cox regression analyses were performed with OS as the primary endpoint and SNP genotype as the only variable. SNP genotype was derived via imputation. All analyses were performed on cases with no missing data. Hazard ratios demonstrate the associated risk with loss of the functional allele for each SNP. Statistical significance was corrected for multiple testing using the Bonferroni correction to P=8.6x10^{-5}. This resulted in no significant results.

Table 3.11. Univariate analyses of response at 12 weeks for *TLR*, *FPR* and *NOD* SNPs

SNP	N	OR (95% CI)	*P*	SNP	N	HR (95% CI)	*P*
TLR5				*TLR10*			
rs2072493	1,488	1.04 (0.84–1.28)	0.71	rs4833095	1,649	1.06 (0.89–1.26)	0.53
rs5744174	1,649	1.08 (0.94–1.24)	0.30	rs5743551	1,613	1.06 (0.89–1.24)	0.50
rs851139	1,538	1.123 (0.97–1.31)	0.13	rs5743557	1,625	1.02 (0.85–1.24)	0.81
IFIH1				rs5743565	1,628	1.04 (0.86–1.26)	0.71
rs1990760	1,649	0.95 (0.83–1.10)	0.50	rs5743592	1,640	1.03 (0.85–1.25)	0.76
rs3747517	1,649	1.05 (0.90–1.10)	0.54	rs5743611	1,604	0.94 (0.72–1.22)	0.63
TLR9				*NOD1*			
rs352140	1,533	1.02 (0.89–1.18)	0.78	rs2075818	1,535	1.02 (0.86–1.21)	0.80
rs352139	1,547	1.01 (0.88–1.16)	0.91	rs2075820	1,533	1.02 (0.86–1.21)	0.82
rs5743836	1,510	0.97 (0.79–1.19)	0.76	*DDX58*			
TLR1				rs10813831	1,649	0.99 (0.84–1.15)	0.86
rs4543123	1,613	1.04 (0.87–1.24)	0.71	rs3205166	1,593	1.03 (0.89–1.20)	0.68
rs4833095	1,649	1.06 (0.89–1.26)	0.53	rs56309110	1,639	0.89 (0.75–1.05)	0.16
rs5743551	1,613	1.06 (0.89–1.27)	0.50	rs669260	1,639	1.12 (0.92–1.37)	0.25
rs5743557	1,625	1.02 (0.85–-1.24)	0.81	rs659527	1,631	1.03 (0.89–1.19)	0.68
rs5743565	1,628	1.04 (0.86–-1.26)	0.71	rs10813821	1,649	0.93 (0.81–1.07)	0.31
rs5743592	1,640	1.03 (0.86–-1.25)	0.76	rs3739674	1,631	1.03 (0.89–1.20)	0.67
rs5743611	1,604	0.94 (0.72–1.22)	0.63	rs11795343	1,631	1.03 (0.90–1.19)	0.65
TLR2				rs1133071	1,639	0.88 (0.75–1.02)	0.09
rs11938228	1,649	1.041 (0.90–1.20)	0.58				

rs1816702	1,649	0.79 (0.64–0.98)	0.04	*NOD2*			
rs1898830	1,614	1.00 (0.87–1.17)	0.95	rs1077861	1,627	1.01 (0.87–1.16)	0.95
rs201786064	1,576	0.89 (0.73–1.10)	0.28	rs2066842	1,627	1.01 (0.87–1.18)	0.91
rs3804099	1,649	0.92 (0.80–1.06)	0.26	rs2066843	1,628	1.01 (0.87–1.18)	0.87
rs3804100	1,649	0.84 (0.64–1.09)	0.18	rs2067085	1,601	0.99 (0.86–1.15)	0.93
rs4696480	1,611	0.94 (0.81–1.08)	0.36	rs2076756	1,649	1.01 (0.87–1.18)	0.89
rs76112010	1,635	1.13 (0.97—1.31)	0.11	rs3135499	1,635	0.95 (0.82–1.09)	0.45
rs7656411	1,625	1.02 (0.87–1.20)	0.84	rs3135500	1,649	0.95 (0.82–1.09)	0.43
rs7696323	1,649	1.14 (0.98–1.32)	0.01	rs5743289	1,649	0.99 (0.83—1.19)	0.94
TLR6				rs8057341	1,524	1.02 (0.87–1.19)	0.86
rs1039559	1,596	1.01 (0.88–1.17)	0.88	*FPR2*			
rs3775073	1,573	1.00 (0.86–1.16)	0.96	rs11666254	1,649	0.97 (0.84–1.11)	0.67
rs5743810	1,649	0.96 (0.83–1.10)	0.54	rs4802859	1,627	0.93 (0.80–1.08)	0.34
rs5743827	1,581	1.01 (0.87–1.18)	0.85	*FPR3*			
rs4543123	1,613	1.04 (0.87–1.24)	0.70	rs17695224	1,649	0.99 (0.85–1.17)	0.91

Additive logistic regression analyses were performed with response at 12 weeks as the primary endpoint and individual SNP genotype as the only variable. SNP genotype information was derived via imputation. All analyses were performed on all cases with no missing data and SNPs with a MAF>5%. Hazard ratios demonstrate the associated risk with loss of the functional allele for each SNP. Statistical significance was corrected for multiple testing using the Bonferroni correction to P=8.6 x10^{-5}. This resulted in no significant results.

3.4 Discussion

3.4.1 Previously explored PRP SNPs and Oxaliplatin Efficacy

Using data COIN and COIN-B, with a sample size of nearly 2,000 patients, there is no evidence to support the detrimental effect of variation in SNPs within three PRPs *FPR1, TLR3* and *TLR4* have on survival and response outcomes in cancer patients receiving oxaliplatin-based chemotherapy. This was supported by a confirmatory cohort (SCOT) which provided no evidence of a significant relationship between these variants and two survival endpoints (OS and DFS) (Iveson et al., 2018). This gives a combined sample size of over 5,000 patients. Each cohort independently has a power to detect associations of similar strengths (HR=1.18–3.23) to those previously reported for rs867228 and rs4986790 (Tesniere et al., 2010, Vacchelli et al., 2016). The fact that these two large, independent studies show similar results, coupled with the relatively small previously reported effect sizes, suggest that it is unlikely that the previously described univariate relationship between SNPs in PRPs and survival outcomes for cytotoxic anti-cancer therapies are true associations (Vacchelli et al., 2015, Castro et al., 2011, Tesniere et al., 2010).

The same investigations showed that some associations between PRPs such as *FPR1* were dependent on the presence of genetic variance of other genes such as *TLR3* and *TLR4* (Vacchelli et al., 2015). However, when examining rs867228 on in both COIN and COIN-B and SCOT, there was no significant association between rs867228 regardless of *TLR3* or *TLR4* variation. In our cohorts, functionality of *TLR3* or *TLR4* has no influence over the significance of a SNP in *FPR1* in relation to survival.

The discordance between results from COIN and COIN-B (which were confirmed in SCOT) and those previously reported may be due to sample size. Statistical studies with smaller sample sizes as previously published (Table 3.1) have a decreased power and an increased likelihood of type I errors (Forstmeier et al., 2017). This has been acknowledged by the investigators exploring the effect of *TLR3* which was one of the larger cohorts to explore this effect (Castro et al., 2011). Furthermore, there was no evidence that these investigations corrected for multiple testing (for example for using

different inheritance models) which can result in false positive results (Rice et al., 2008).

Similar difficulty in validating previously reported possible predictive biomarkers for CRC treatment with platinum-based therapies have been seen. The prospective trial, MAVERICC, specifically designed to support preliminary data surround the role of *Excision Repair Cross-Complementation Group 1 (ERCC1)* expression in platinum-based therapies, showed *ERCC1* levels in mCRC patients treated with mFOLFOX6-bevacizumab was not associated with response to oxaliplatin (Park et al., 2001, Park et al., 2003, Parikh et al., 2019).

Whilst these results do not dismiss the suggestions that PRPs may still have a modulatory effect on the immune system in general, it does imply that these specific variants within *FPR1*, *TLR3* and *TLR4* should not be used as predictive biomarkers for CRC moving forwards. They also highlight the need for validation of suggestive findings in small-to medium-sized studies by large independent cohorts to either support or oppose the significant associations. It has also been argued as of a result of this and similar studies that the lack of such validation should result in caution and the avoidance of generalisation when discussing positive preliminary genetic data (Battaglin and Lenz, 2019). Whilst this does not mean that pharmacogenetic studies should not be performed, they should perhaps move away from single-gene investigations on heterogeneous retrospective cohorts.

To maximise the impact of pharmacogenetic studies in the future, it is important to move towards applying statistical techniques, like those in this chapter and the studies showing an association with PRPs and treatment response, on high-quality cohorts. This includes a move towards homogenous cohorts with tumours of similar molecular subtypes, large patient cohorts for meta-analysis and specifically designed prospective studies. Additionally, whole genome (including the cancer genome) analyses to take into account germline and tumour interactions may increase the relevance of the results, which will be performed in other chapters of this thesis.

3.4.2 Extended Profile of PRPs

When widening this study to include variants from genes within the same (FPRs and TLRs) and closely related (RIG-I-Type Receptors and Nucleotide-binding oligomerisation domain-containing proteins) families, there were some variants which had a nominal association with OS and response (P<0.05). However it was important to ensure that the significance level was stringent and adjusted to reduce the likelihood of false discovery (Forstmeier et al., 2017). When a correction for multiple testing was applied, there is no evidence that these SNPs had an effect on OS or response. This supports the conclusions from the attempts to validate the effect of genetic variation in *FPR1, TLR3* and *TLR4* – it is unlikely that PRPs will be clinically useful predictive biomarkers in early (SCOT) and late-stage (COIN and COIN-B) CRC patients (Gray et al., 2019).

3.4.3 Strengths and weaknesses

Strengths of this investigation into genetic variation within three PRPs (*FPR1*, *TLR3* and *TLR4*) include the large sample size (COIN and COIN-B n=1,948 patients) compared to original studies and inclusion of additional variants known to determine clinical practice including age and tumour biomarker mutation status for multivariate analyses (Vacchelli et al., 2015, Castro et al., 2011, Tesniere et al., 2010). Additionally, the inclusion of a confirmatory cohort (SCOT, n=2,929 patients) of patients with Stage II and III CRC which also demonstrates that these variants do not have a predictive effect in multiple disease stages (Iveson et al., 2018).

The lack of complete data for known prognostic variables such as MSI may reduce the clinical relevance of this attempted validation (Popat et al., 2005). Additionally, the multivariate analysis in the confirmatory cohort did not include the same covariates and this should be acknowledged when comparing results between COIN and COIN-B and SCOT. This highlights the importance of further prospective investigations specifically designed to investigate the predictive nature of biomarkers that may be carried forward to a clinical setting for any disease. However, exploitation of already

established rich datasets, such as COIN and COIN-B and SCOT, where they are available is a useful starting point for external validation.

Investigation of the extended profile of PRP variants for associations (Section 3.9) allowed for a more detailed exploration of the role of this family of proteins in oxaliplatin-based chemotherapy treatment outcomes. It does, however, have some drawbacks. Due to lack of sequencing of the sex chromosomes during genotyping, members of the Toll Like Receptor family were excluded due to their genomic location on the X chromosome. Furthermore, both the survival and response analyses were performed with a univariable regression analysis under one inheritance model and therefore not as comprehensive as the earlier studies. However, this did reduce the number of independent tests and the likelihood of type I errors (Finner and Roters, 2002). Additionally, at this time the novel, negative findings of the extended profile of PRPs (Section 3.9) have not been validated in an independent cohort. This is despite the findings of initial investigation into variations in *FPR1*, *TLR3* and *TLR4* (Sections 3.5 and 3.6) and Battaglin and Lenz (2019) highlighting the importance of independent validation of potential biomarkers for CRC treatment outcomes.

3.4.4 Overall conclusions and follow up studies

Whilst it cannot be denied that the immune system plays an important role in the activity of chemotherapy, including oxaliplatin-based chemotherapies, this investigation has been unable to provide evidence to support the previously reported associations between SNPs in PRPs and oxaliplatin-based chemotherapy treatment outcomes (Vacchelli et al., 2015, Tesniere et al., 2010, Castro et al., 2011, Gray et al., 2019). An extended profile of additional SNPs across a number of chromosomes was also unable to support the notion that variations in the PRPs has an impact on prognosis in CRC patients (Grizzi et al., 2018, Markman and Shiao, 2015).

As the findings of the extended profile have yet to be validated, an important future aim on this investigation is to validate the negative findings in independent cohorts. This may include SCOT, which can provide information on survival in patients with

earlier stage CRC and other medium to large clinical cohorts with more complete data on response to oxaliplatin-based chemotherapies (Iveson et al., 2018).

The findings of this chapter do not dismiss the role that the innate immune system may play in CRC outcomes including OS and radiological response. However, further investigations, particularly in molecularly similar cohorts, is required to fully explore the relationship between PRPs, the innate system, and the activity and outcomes of cytotoxic anti-cancer treatments in CRC patients.

4. Chapter 4: Genome wide search for common germline variants that influence response to oxaliplatin-based chemotherapy

4.1 Introduction

4.1.1 Oxaliplatin as treatment for advanced colorectal cancer

Oxaliplatin alone has been shown to have modest activity in mCRC patients, with reported response rates of <25% (Machover et al., 1996, Zori Comba et al., 2001, Diaz-Rubio et al., 1998). The addition of 5-FU to oxaliplatin increased response rates to over 50%, as a result of the interaction between the two molecules (de Gramont et al., 2000, Cunningham et al., 2009). Capecitabine, which when metabolised mimics continuous infusion of 5-FU, can be taken orally and has been shown to have similar efficacy to 5-FU (Chintala et al., 2011, Van Cutsem et al., 2004, Johnston and Kaye, 2001). This research has resulted in oxaliplatin paired with either 5-FU or capecitabine being used as adjuvant and neoadjuvant treatments for CRC in patients at different disease stages (Boland and Fakih, 2014, Sobrero et al., 2018, Coutinho and Grothey, 2016).

Whilst oxaliplatin and other platinum-based therapies have been shown to be efficient in the treatment of CRC, they have a number of serious side effects particularly in the gastrointestinal, hemopoietic, and peripheral nervous systems (Alcindor and Beauger, 2011). There is also some evidence that this treatment may impact on fertility in younger patients of both sexes (Hrushesky et al., 1999, Cercek et al., 2013). Oxaliplatin-induced neurotoxicity is the most common dose dependent severe toxicity that causes dose reduction, delay or cessation of treatment (Kiernan and Krishnan, 2006, Argyriou, 2015, Kidwell et al., 2012).

By predicting if patients are likely to respond to oxaliplatin-based chemotherapy, clinicians may be able explore other treatment options which do not include oxaliplatin, such as FOLFIRI (Kamnerdsupaphon et al., 2007). This will allow for more effective shared decision-making (SDM) between patients and clinicians which has been shown

to improve patient satisfaction and quality of life, which is especially important in late-stage cancer patients (Kashaf and McGill, 2015, Nayak et al., 2017).

4.1.2 Oxaliplatin in the COIN and COIN-B trials

FOLFOX (5-FU) and XELOX (capecitabine) was given to patients in the COIN and COIN-B trials and are still used to-date in clinical settings for aCRC patients (National Institute for Health and Care Excellence, 2020a, Cassidy et al., 2011, Maughan et al., 2011, Adams et al., 2011). Traditionally, patients receive FOLFOX; however, there is evidence of improved quality of life and lack of inferiority of outcomes in patients who are given oral XELOX. This has resulted in a rise of its use in clinical settings (Conroy et al., 2010, Cassidy et al., 2011, Guo et al., 2016). Present recommendations state that patients with anaemia, bleeding disorders and gastrointestinal issues are given FOLFOX, with those who are elderly, have diabetes or immunodeficiency given XELOX (Guo et al., 2016). In COIN, two thirds of patients chose to receive oral capecitabine and oxaliplatin in the form of XELOX, reflecting the increasing incidence of this treatment in the clinical settings, especially in late-stage cancers (Adams et al., 2011). This use of both treatments is encouraged in the NICE guidelines for treatment of aCRC (National Institute for Health and Care Excellence, 2020a).

4.1.3. Known biomarkers for response to oxaliplatin-based chemotherapies in colorectal cancer

Many recent studies have focused on the use of tumour biomarkers to guide both new, targeted treatments for CRC and more established treatments such as oxaliplatin-based chemotherapy (Febbo et al., 2011). It has been shown that cancers with a high frequency of microsatellite instability (MSI-H) have resistance or poor response to 5-FU, when delivered in conjunction with oxaliplatin (Jo and Carethers, 2006). There is also evidence that mutations in genes involved in the MAPK (ERK) pathway, which are known to be prognostic for survival, play a key role in response to oxaliplatin (Fang and Richardson, 2005). In particular, patients with *BRAF* or *KRAS* mutant tumours are more likely to have poor response to many types of treatment including chemotherapy

as a first-line treatment or chemotherapy coupled with other treatments such as monoclonal antibodies (Garcia-Carbonero et al., 2020, Basso et al., 2013, Scartozzi et al., 2015).

Although tumour mutation status plays a key role in how a patient responds to treatment, it is important to note that treatment response is also due to the activity of germline cells such as those in the adaptive immune system (Galon et al., 2006) (Chapter 1, Section 3.2). Therefore, the exploration of germline biomarkers for response is also important for reliable prediction of this phenotype.

The key findings associated with PFS (an indirect measure of response) in CRC (Chapter 1, Section 3) include polymorphisms within genes associated with DNA repair (including the *Xeroderma pigmentosum* family of genes; *Ataxia Telangiectasis Mutated (ATM)* and *Excision Repair Cross-Complementation Group 5 (ERCC5)*) which have been found in patients from different racial backgrounds treated with oxaliplatin-based chemotherapies (Hu et al., 2019, Kweekel et al., 2009). Fewer studies have used response as the detectable outcome. However, one that has employed this method has detected associations for variants in the promoter region of *ERCC5*, another DNA repair protein (Chen et al., 2009a). This is likely due to oxaliplatin's mechanism of action – DNA damage and triggering of apoptosis through inhibition of DNA and RNA synthesis (Alcindor and Beauger, 2011).

4.1.4 Hypothesis and aims

The main hypothesis of this chapter is that there are germline genetic variants associated with response to oxaliplatin. The specific aims are:

- Search for germline SNPs statistically associated with response to oxaliplatin.
    - Perform a univariate GWAS for response on all genotyped patients from COIN and COIN-B.
    - Perform a multivariate GWAS for response on the same patients to include known prognostic factors.
- Perform survival analyses on SNPs suggestive of association with response to oxaliplatin.
- Determine if there is any supporting evidence for the newly discovered SNPs.

4.2 Methods

4.2.1 Patient inclusion and endpoints

All patients in COIN and COIN-B were given oxaliplatin-based chemotherapies, either FOLFOX or XELOX as part of this trial. However, due to changes in chemotherapy delivery (continuous vs intermittent), treatment was only comparable for the first 12 weeks and response at this time was used as the variable for this investigation. Therefore, of the 2,671 patients recruited, 1,649 had complete SNP genotypes and data for response at 12 weeks. Due to the homogeneity of response outcomes regardless of chemotherapy type or if patients received cetuximab (Chapter 2.3.3), all patients were combined for analyses.

4.2.2 Statistical analyses

4.2.2.1 Genome wide association studies

Details of methodology for univariate and multivariate additive logistic GWAS using Plink version 1.9 can be found in Chapter 2, Section 4.3 (Purcell et al., 2007). Only SNPs with a MAF ≥5% were included in this investigation.

Additive logistic regressions to test for association between known covariates and response were carried out using the -no-snp function and are reported in Table 4.1. These covariates were also included in the multivariate GWAS.

Odds ratios reported in Table 4.2 are based on two-sided genome wide and suggestive significance thresholds (α), a power (1-β) of 80%, response rate of 58% (response rate within COIN and COIN-B) and MAF ≥5%. These were calculated using the *genpwr* package in R-version 3.5.2 (Moore and Jacobson, 2020, Hong and Park, 2012, R Core Team, 2018).

Table 4.1. Factors included in the multivariate analyses of response to chemotherapy and overall survival

	Response (n=1332) OR (95% CI)	***P***	**OS (n=1332) HR (95% CI)**	***P***
Age	1.01 (0.99–1.02)	0.32	1.00 (1.00–1.01)	0.30
Sex				
Male	1.0		1.0	
Female	0.96 (0.75–1.22)	0.73	1.03 (0.90–1.17)	0.70
Disease site				
Colon	1.0		1.0	
Rectum	0.96 (0.75–1.22)	0.74	0.85 (0.74–0.98)	0.03
WHO performance status				
0–1	1.0		1.0	
2	0.69 (0.43–1.08)	0.10	1.43 (1.10–1.85)	6.8×10^{-3}
Primary tumour resected				
No	1.0		1.0	
Yes	0.82 (0.64–1.03)	0.10	0.71 (0.62–0.81)	7.2×10^{-7}
Local recurrence	0.59 (0.34–1.02)	0.06	0.75 (0.54–1.01)	0.07
White cell count				
<10,000/mcL	1.0		1.0	
≥10,000/mcL	0.98 (0.95–1.01)	0.29	1.64 (1.43–1.87)	4.0×10^{-13}
KRAS mutation status				
Wild-type	1.0		1.0	
Mutant	0.47 (0.37–0.60)	1.2×10^{-9}	1.51 (1.33–1.72)	1.9×10^{-10}
NRAS mutation status				
Wild-type	1.0		1.0	
Mutant	0.50 (0.29–0.86)	0.01	1.55 (1.17–2.06)	2.3×10^{-3}
BRAF mutation status				
Wild-type	1.0		1.0	
Mutant	0.36 (0.24–0.55)	1.5×10^{-6}	2.43 (1.98–2.97)	2.0×10^{-16}
Cetuximab treatment				
No	1.0		1.0	
Yes	1.09 (0.86–1.38)	0.47	1.05 (0.92–1.20)	0.45
Chemotherapy regimen				
XELOX	1.0		1.0	
FOLFOX	1.13 (0.89–1.43)	0.32	0.95 (0.84–1.07)	0.41

Multivariate logistic regression analyses (for response: responders – complete or partial response – vs non-responders – stable or progressive diseases) and multivariate cox regression (for OS) of prognostic factors were performed. Analysis excludes cases with missing values for any prognostic factor or response data at 12 weeks. Odds and Hazard Ratios for each variable were adjusted for all other covariates listed. Positive Odds Ratios indicate an increased likelihood of response and positive Hazard Ratios indicate an increased likelihood of early death. Odds and Hazard Ratios for reference groups are listed as 1.0.

Table 4.2. Detectable Odds Ratios at 80% power in logistic regression analysis with response at 12 weeks

Test	Cases	Responders	*P*	Detectable OR
Univariate additive GWAS				
	1,649	954	$1.0x10^{-5}$	2.54
	1,649	954	$5.0x10^{-8}$	3.16
Multivariate additive GWAS				
	1,332	788	$1.0x10^{-5}$	2.91
	1,332	788	$5.0x10^{-8}$	3.79

Detectable hazard ratios (ORs) for each SNP takes into account the minimum MAF (5%), sample size, probability of failure and adjusted two-sided (α) suggestive significance ($1.0x10^{-5}$) and genome wide significance ($5.0x10^{-8}$). All calculations were based on a power (1–β) of 80%.

4.2.2.2 Survival analyses

Lead SNPs from all independent loci suggestive of association with response to chemotherapy were tested for associations with OS through additive univariate and multivariate Cox regressions using the Survival package in R-version 3.5.2 (Therneau, 2020, R Core Team, 2018). Significance was adjusted for multiple testing using Bonferroni correction (Bland and Altman, 1995). Additive Cox regressions to test for associations with covariates was performed in the same manner, but without the germline SNP data (Table 4.1).

4.2.2.3 Comparing results from subsets

To test for differences in association for individual SNPs between subgroups separated by treatment, the following calculation was used:

$$P_{Differnece}\frac{(Group\ 1\ \beta - Group\ 2\ \beta)^2}{((Group\ 1\ SE)^2 - (Group\ 2\ SE)^2\ )}$$

Where β is the log(Odds Ratio) for the SNP of interest and SE is the standard error for that same output.

This is based on a chi-squared test with one degree of freedom and therefore a *P*-value for the difference between the associations of these two groups. This was held to the standard significance threshold of $P<0.05$ with no adjustment required for multiple testing.

4.2.2.4 Gene-based and gene-set analyses

Gene-based and gene-set analyses were performed in MAGMA using standard protocols (de Leeuw et al., 2015). A competitive model was used for gene set analyses with outputs being adjusted to take into account false discovery rate (FDR) (Chapter 2, Section 4.2.2).

4.2.3 Other bioinformatic analyses

Regional association analysis for suggestive SNPs were visualised using LocusZoom (Pruim et al., 2010). eQTL information for these same SNPs was sourced from the GTEx Project Database (Chapter 2, Section 2).

4.3 Results

4.3.1 Statistical Power

Based on a sample size of 1,649 patients, the univariate GWAS analysis had 80% power to detect variants with ORs ≥3.2 at genome wide significance and ≥2.6 at suggestive significance. The multivariate analysis (n=1,332 patients) had 80% power to detect variants with ORs ≥3.8 at genome wide significance and ≥2.9 at suggestive significance (Table 4.2).

4.3.2 Known prognostic factors

Association tests (multilinear logistic and Cox regressions) for known prognostic factors were performed in order to ensure that multivariate analyses included key prognostic factors. Mutations in *KRAS* (OR=0.47, 95% CI=0.37–0.67, $P=1.2\text{x}10^{-9}$), *NRAS* (OR=0.50, 95% CI=0.29–0.86, *P*=0.01) and *BRAF* (OR=0.36, 95% CI=0.24–0.55, $P=1.5\ \text{x}10^{-6}$) mutation status were significantly associated with worse response at 12 weeks. Disease site (HR=0.85, 95% CI=0.74–0.98, *P*=0.03), WHO performance status (HR=1.43, 95% CI=1.10–1.85, $P=6.8\text{x}10^{-3}$), primary tumour resection status (resected HR=0.71, 95% CI=0.62–0.81, $P=7.2\text{x}10^{-7}$), white cell count (HR=1.64, 95% CI=1.43–1.87, $P=4.0\text{x}10^{-13}$), *KRAS* (HR=1.51, 95% CI=1.33–1.72, $P=1.9\text{x}10^{-10}$), *NRAS* (HR=1.55, 95% CI=1.17–2.06, $P=2.3\text{x}10^{-3}$) and *BRAF* (HR=2.43, 95% CI=1.98–2.97, $P=2.0\text{x}10^{-16}$) mutation status were significantly associated with OS (Table 4.1).

4.3.3 Genomic inflation and population substructure

The distribution of expected and observed *P*-values for the univariate and multivariate GWAS (Figure 4.1A and B) and the genomic inflation factor for both analyses (λ=1.01) demonstrates there is no underlying abnormal population substructure.

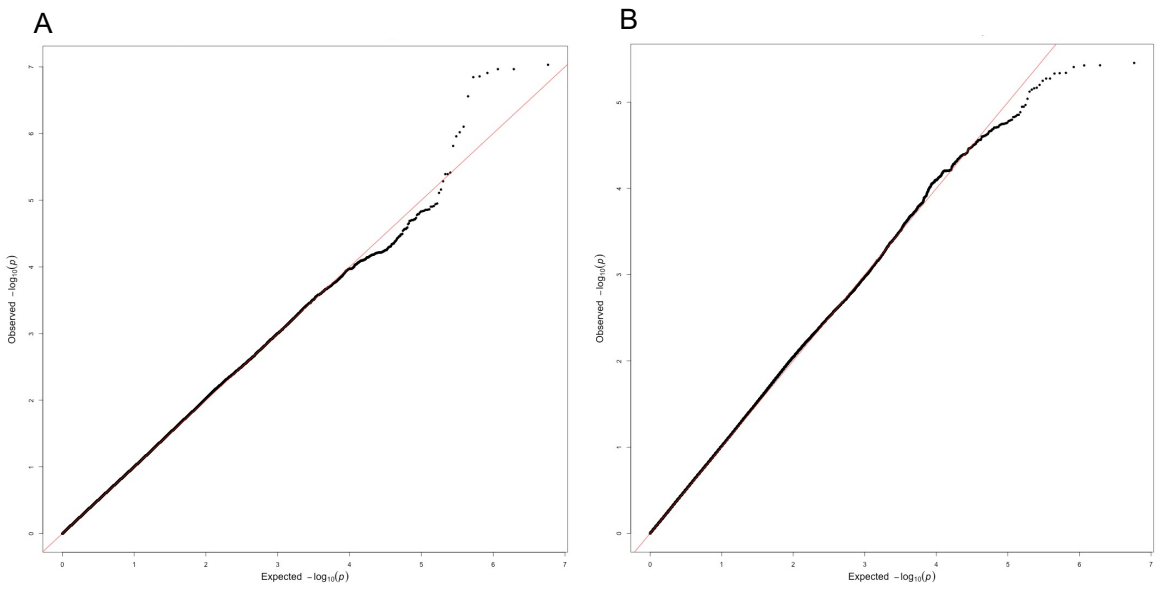

Figure 4.1. Q-Q plots of observed vs expected P values for response to chemotherapy in (A) univariate additive model (n=1,649 patients) and (B) multivariate additive model (n=1,332 patients)
Genomic inflation factor (λ) for both univariate and multivariate models=1.01.

4.3.4 Univariate analysis of response to chemotherapy

No SNPs were associated with response to chemotherapy at genome-wide significance. Seventeen variants at four independent loci were found to be suggestive of association (Figure 4.2, Table 4.3). The most significant locus was in cytoband 10p15.3, with eleven SNPs of suggestive association found in LD (lead SNP rs10903369, OR=2.11, P=9.3x10^{-8}).

4.3.5 Multivariate analysis of response to chemotherapy

When taking into account prognostic factors (Table 4.1), there were no variants associated with response to oxaliplatin at genome-wide significance. Sixteen variants at five independent loci were found to be suggestive of association (Figure 4.3, Table 4.4). The most significant locus was found at 5q14.1; lead SNP rs4704514 (OR=1.54, P=3.5x10^{-6}).

4.3.6 Loci suggestive of significance

Of the four loci suggestive of association with response to oxaliplatin under univariate analyses, two lead SNPs have genes in eQTL (Table 4.5) – rs10903369 in an eQTL for *WDR37* and *IDI1* and rs3759992 is an eQTL for *NPIPB4*, *RP11-645C24.5* and *RRN3P1*.

Of the five suggestive loci under multivariate analyses, three lead SNPs have genes in eQTL (Table 4.5), rs4704514 is an eQTL for *SCAMP1-AS1*, rs2086382 is an eQTL for *WDR37* and rs6585148 is an eQTL for *ACSL5*, *RP11-324O2.3* and *GPAM*.

The only locus suggestive of significance in both univariate and multivariate analysis was in cytoband 10p15.3. The lead SNP at this locus in the multivariate analysis was rs2086382 (OR=2.12, P=3.7x10^{-6}), the lead SNP in the univariate analysis was rs10903369 (OR=2.11, P=9.3x10^{-8}).

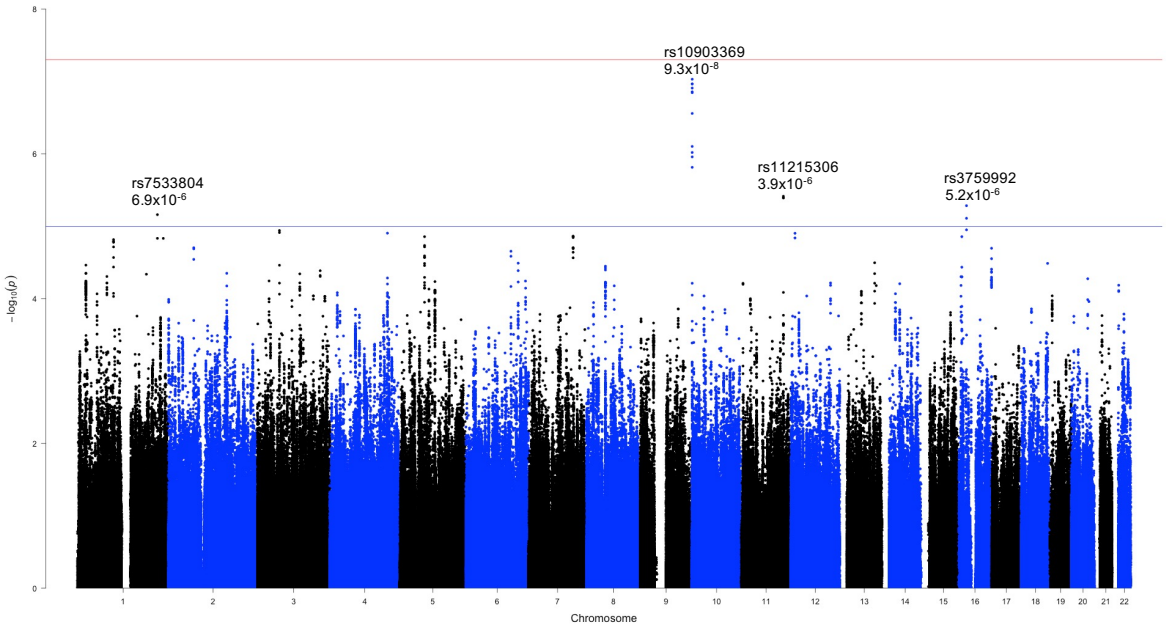

Figure 4.2. Manhattan plot showing univariate GWAS for response to chemotherapy (n=1,649 patients)

Responsive patients had complete or partial response. Non-responsive patients were those with stable or progressive disease as defined by RECIST 1.0 guidelines. MAF>0.05. Genome wide significance (red line) $P<5.0x10^{-8}$. Suggestive association (blue line) $P<1.0x10^{-5}$.

Table 4.3. Common variants suggestive of association with response to chemotherapy under univariate analyses ($P<1.0x10^{-5}$)

Cytoband	SNP	MAF (%)	Alleles REF/ALT	OR (95% CI)	*P*
10p15.3	rs10903369	8.8	T/C	2.11 (1.61–2.78)	$9.3x10^{-8}$
11q23.3	rs11215306	15.5	C/T	0.64 (0.53–0.77)	$3.9x10^{-6}$
16p12.2	rs3759992	21.0	A/G	0.68 (0.58–0.80)	$5.2x10^{-6}$
1q41	rs7533804	6.9	C/T	0.53 (0.41–0.70)	$6.9x10^{-6}$

Lead SNPs of independent loci shown. BP=Base Position, MAF=Minor allele frequency, REF=Reference allele, ALT=Alternative allele, OR=Odds ratio, CI=Confidence interval. *P*=*P*-value.

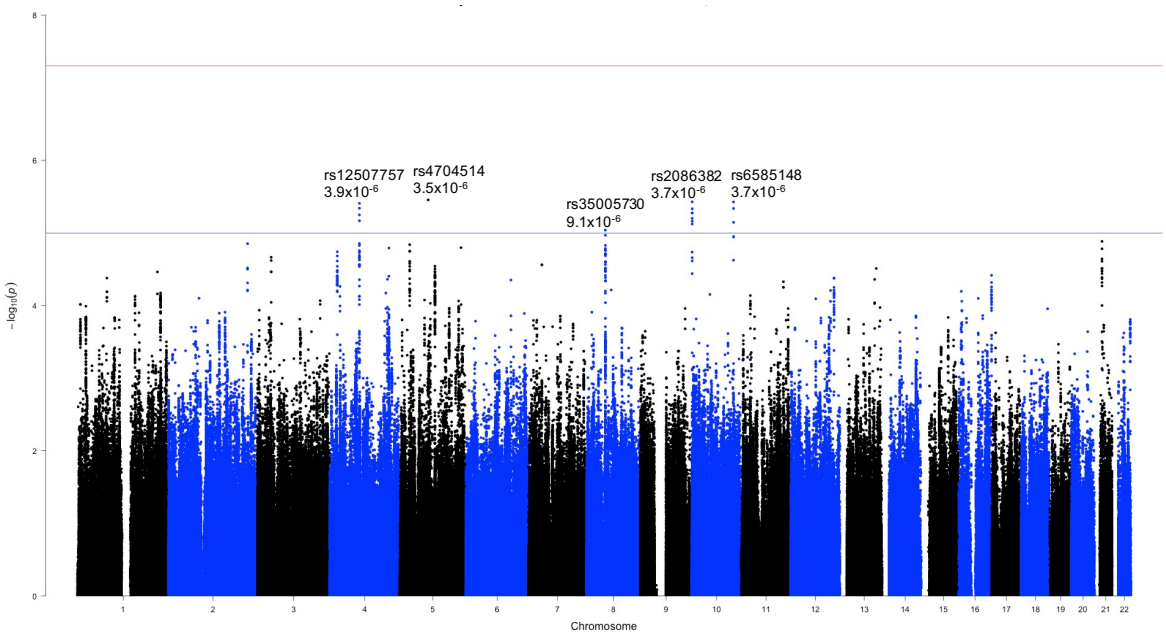

Figure 4.3. Manhattan plot showing multivariate GWAS results for response to chemotherapy (n=1,332 patients)

Responsive patients had complete or partial response. Non-responsive patients were those with stable or progressive disease as defined by REICIST 1.0 guidelines. MAF>0.05. Genome-wide significance (red line) $P<5.0x10^{-8}$. Suggestive association (blue line) $P<1.0x10^{-5}$.

Table 4.4. Common variants suggestive of association with response in multivariate analyses

Cytoband	SNP	MAF (%)	Alleles REF/ALT	OR (95% CI)	*P*
5q14.1	rs4704514	28.0	C/T	1.54 (1.28–1.85)	$3.5x10^{-6}$
10p15.3	rs2086382*	8.9	G/A	2.12 (1.54–2.91)	$3.7x10^{-6}$
10q25.2	rs6585148	42.0	C/A	1.47 (1.25–1.74)	$3.7x10^{-6}$
4q21.21	rs12507757	41.1	T/C	1.50 (1.26–1.78)	$3.9x10^{-6}$
8q11.21	rs35005730	15.0	A/-	1.70 (1.35–2.16)	$9.1x10^{-6}$

Lead SNPs for each independent loci shown. Odds Ratios are adjusted for age, sex, disease site (colon vs rectum), World Health Organization (WHO) performance status (0 or 1 vs 2), primary tumour resection (unresected vs resected vs local recurrence), tumour *KRAS*, *NRAS*, and *BRAF* mutation status (mutated vs wild-type), patient white blood cell count (<10 000 cells/μL vs ≥10 000 cells/μL), addition of cetuximab (yes vs no) and chemotherapy regimen (XELOX vs FOLFOX). *=locus suggestive of significance in univariate analysis – lead SNP from univariate analysis=rs10903369. BP=Base Position, REF=Reference allele, ALT=Alternative allele, -=deletion, OR=Odds ratio, CI=Confidence interval. *P*=*P* value.

Table 4.5. Genes associated with significant loci from multivariate GWAS

Cytoband	SNP	Nearby Genes	Genes in eQTL
Univariate			
10p15.3	rs10903369	*WDR37, IDI1*, *IDI2*, *GTPBP, ADARB2*, *LARP4B*, *C10orf110*	*WDR37, IDI1*
11q23.3	rs11215306	*CADM1*, *FAM55B*, *FAM55D*	
16p12.2	rs3759992	*METTL9*, *OTOA*, *RRN3P1*, *SLC7A5P2, IGSF6, NPIPL3*, *UQCRC2*	*NPIPB4*, *RP11-645C24.5*, *RRN3P1*
1q41	rs7533804	*TFGB2*, *RRP15*	
Multivariate			
5q14.1	rs4704514	*LHFPL2*, *SCAMP1*, *ARSB*, *AP3B1*	*SCAMP1-AS1*
10p15.3	rs2086382	*WDR37, IDI1*, *IDI2*, *GTPBP, ADARB2*, *LARP4B*, *C10orf110*	*WDR37*
10q25.2	rs6585148	*TECTB*, *GUCY2G*, *ACSL5*, *GPAM*, *ZDHHC6*, *VTI1A*	*ACSL5*, *RP11-324O2.3*, *GPAM*
4q21.21	rs12507757	*GDEP*, *ANTXR2*, *PRDM8*	
8q11.21	rs35005730	*SNTG1*	

4.3.6.1 Effect of SNPs suggestive of association with response on overall survival

Of the four loci suggestive of association to response in the univariate analysis, only rs10903369 had a significant association with OS (HR=0.83, *P*=0.01, Table 4.6) Individuals with the CC genotype (blue curve in Figure 4.4) have a median survival time 3.8 months longer than individuals with the TT genotype. The increased survival time for the minor allele of rs10903369 is also supported by the increased rate of response for the same allele (Table 4.7). The lead SNP from the same locus in a multivariate model (rs2086382) was also significantly associated with OS (HR=0.77, $P=3.0\text{x}10^{-3}$, Table 4.6). Moving forwards, only rs10903369 was analysed from this locus to eliminate multiple testing.

4.3.6.2 Effects of covariates

There was no significant difference between the effect size (OR and HR) or direction of effect for the minor allele of rs10903369 in the univariate and multivariate models for response or OS (Table 4.8), suggesting that the covariates have limited influence on this SNP.

4.3.6.3 Associations between rs10903369 and response and overall survival and different treatments

The group with the most significant associations between rs10903369 and response was those who were administered XELOX (Univariate OR=2.19, $P=6.4\text{x}10^{-6}$; Multivariate OR=2.22, $P=6.4\text{x}10^{-5}$). There was no significant difference in the effect of this SNP between patients who receive XELOX and FOLFOX in either model. In contrast, the most significant association for OS was found in patients who were given FOLOX under a multivariate model (HR=0.66, 95% CI=0.48–0.90, $P=9.4\text{x}10^{-3}$). However, there was only a significant difference in effect size between patients on different chemotherapy types under a multivariate model, however this did not take into account multiple testing (*P*=0.03, Table 4.8).

Table 4.6. Suggestive SNPs for response at 12 weeks and their effect on overall survival

Cytoband	SNP	Alleles REF/ALT	HR (95% CI)	*P*
Univariate				
10p15.3	rs10903369	T/C	0.83 (0.72–0.96)	0.01*
11q23.3	rs11215306	C/T	1.01 (0.91–1.13)	0.83
16p12.2	rs3759992	A/G	0.97 (0.88–1.08)	0.56
1q41	rs7533804	C/T	1.00 (0.85–1.17)	0.97
Multivariate				
5q14.1	rs4704514	C/T	0.89 (0.80–0.98)	0.02*
10p15.3	rs2086382[#]	G/A	0.77 (0.65–0.92)	$3.0x10^{-3}$
10q25.2	rs6585148	C/A	0.97 (0.88–1.06)	0.46
4q21.21	rs12507757	T/C	0.96 (0.87–1.06)	0.38
8q11.21	rs35005730	A/-	0.99 (0.87–1.12)	0.87

Univariate additive regression analyses were performed with OS as the primary endpoint and SNP genotype as the only variable. Hazard ratios in the multivariate analyses are adjusted for age, sex, disease site (colon vs rectum), World Health Organization (WHO) performance status (0 or 1 vs 2), primary tumour resection (unresected vs resected vs local recurrence), tumour *KRAS*, *NRAS* and *BRAF* mutation status (wild-type vs mutant), patient white blood cell count (<10 000 cells/μL vs ≥10 000 cells/μL), addition of cetuximab (yes vs no) and chemotherapy regimen (XELOX vs FOLFOX). [#]=locus suggestive of significance in univariate analysis - lead SNP from univariate analysis=rs10903369. BP=base position, HR=hazard ratio, CI=Confidence interval. *P*=*P*-value. Analysis was performed on cases with no missing data and individuals who had complete data for response at 12 weeks. SNPs ordered by their statistical significance in relation to response at 12 weeks. Nominally significant ($P<0.05$) results indicated by a star (*) with significant results after Bonferroni correction (threshold $P<5.6x10^{-3}$) highlighted in yellow.

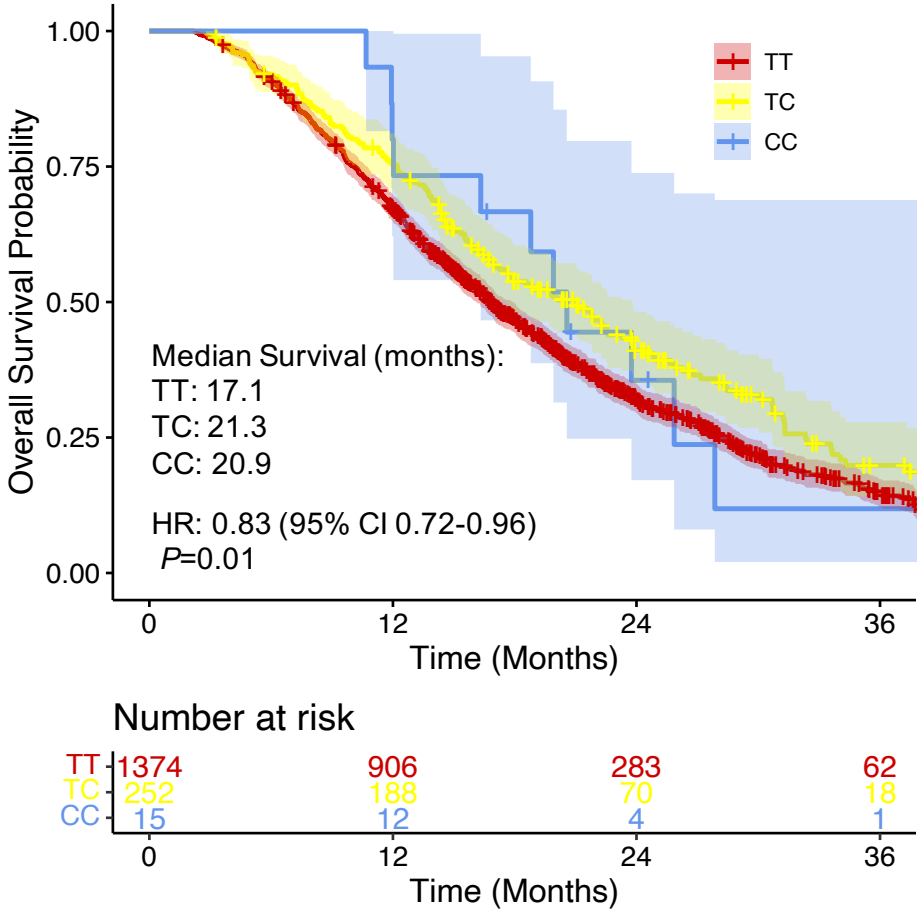

Figure 4.4. Kaplan-Meier curves for overall survival based on rs10903369 genotype

Analysis included all patients with complete response, survival and genotype data for rs10903369 (n=1,641). *P*-values show difference between treatment groups, with shaded areas representing 95% CIs.

Table 4.7. Response rates by rs10903369 genotype

	Response rates by genotype (%)		
	TT	**TC**	**CC**
rs12054810	55	69	93

Table 4.8. Associations between rs10903369 and response at 12 weeks and survival based on treatment regimen

Test	Model	Sample (n)	OR / HR (95% CI)	*P*	$P_{Difference}$
All patients					
Response	Univariate	All patients (1,641)	2.11 (1.61–2.78)	$9.3x10^{-8}$	
	Multivariate	All patients (1,326)	2.10 (1.52–2.89)	$6.9x10^{-6}$	0.86
Survival	Univariate	All patients (1,641)	0.83 (0.72–0.96)	0.01	
	Multivariate	All patients (1,326)	0.79 (0.68–0.91)	$3.0x10^{-3}$	0.39
Chemotherapy type					
Response	Univariate	XELOX only (989)	2.19 (1.56–3.07)	$6.4x10^{-6}$	
		FOLFOX only (652)	2.02 (1.27–3.23)	$3.1x10^{-3}$	0.61
	Multivariate	XELOX only (820)	2.22 (1.50–3.30)	$6.4x10^{-5}$	
		FOLFOX only (508)	1.72 (0.97–3.08)	0.06	0.19
Survival	Univariate	XELOX only (989)	0.81 (0.68–0.98)	0.03	
		FOLFOX only (652)	0.85 (0.67–1.09)	0.20	0.60
	Multivariate	XELOX only (820)	0.83 (0.68–1.02)	0.08	
		FOLFOX only (506)	0.66 (0.48–0.90)	$9.4x10^{-3}$	0.03
Cetuximab					
Response	Univariate	No Cetuximab (992)	1.97 (1.39–2.78)	$6.6x10^{-3}$	
		Cetuximab (649)	2.36 (1.51–3.72)	$1.9x10^{-4}$	0.29
	Multivariate	No Cetuximab (806)	1.90 (1.28–2.81)	$1.3x10^{-3}$	
		Cetuximab (520)	2.42 (1.35–4.36)	$3.2x10^{-3}$	0.21
Survival	Univariate	No Cetuximab (992)	0.81 (0.67–0.97)	0.03	
		Cetuximab (649)	0.85 (0.68–1.07)	0.16	0.61
	Multivariate	No Cetuximab (806)	0.80 (0.65–0.99)	0.04	
		Cetuximab (520)	0.69 (0.51–0.92)	0.01	0.17

Additive univariate and multivariate logistic regressions were performed with response at 12 weeks. Additive univariate and multivariate Cox regression analyses were performed with OS. ORs and HRs are adjusted for age, sex, disease site (colon vs rectum), World Health Organization (WHO) performance status (0 or 1 vs 2), primary tumour resection (unresected vs resected vs local recurrence), tumour *KRAS*, *NRAS*, and *BRAF* mutation status (wild-type vs mutant), patient white blood cell count (<10 000 cells/μL vs ≥10 000 cells/μL) and either addition of cetuximab (yes vs no) or chemotherapy regimen (XELOX vs FOLFOX). Analyses were performed on cases with no missing data and individuals who had complete data for response at 12 weeks. $P_{Difference}$ tests difference in association between two groups based on their beta and standard error.

There was also no significant difference in the effect of rs10903369 on response or OS between patients who did and did not receive cetuximab (Table 4.8).

4.3.6.4 Genomic location and expression of rs10903369

rs10903369 and the associated SNPs in LD are in intron 11 of the gene *WD repeat-containing protein 37 (WDR37)* (Figure 4.5).

In addition, there are three protein coding genes within 150kb upstream of rs10903369, *GTP Binding Protein 4 (GTPBP4), Isopentenyl-Disphosphate Delta Isomerase 1 (IDI1) and Isopentyl-Disphosphate Delta 2 (IDI2).* One RNA coding gene was also upstream, *c10orf110,* also known as *IDI2 Antisense RNA 1 (IDI2-AS1).* There is one protein-coding gene within 150kb downstream, *Adenosine Deaminase RNA Specific B2 (intactive) (ADARB2)* and one non-coding RNA transcript – *NCRNA002200* (Figure 4.4).

rs10903369 was shown to be an eQTL for two genes, both within this region, *WDR37* and *IDI1*. rs10903369 was an eQTL for *WDR37* in the testes (P=9.5x10^{-6}, Figure 4.5A) and for *IDI1* in the Brain-Cortex (P=1.2x10^{-5}, Figure 4.6B) after correcting for multiple testing. Both genes are downregulated by the minor allele of rs10903369 (Figure 4.7).

4.3.7 Gene-based and gene-set analyses

After correcting for multiple testing, no genes were significantly associated with response to chemotherapy when using results from a univariate or multivariate model (Figure 4.6, Table 4.9). A notable association is *WDR37* where rs10903369 is situated. There are also no gene-sets significantly associated with response in either a univariate or multivariate model (Table 4.10).

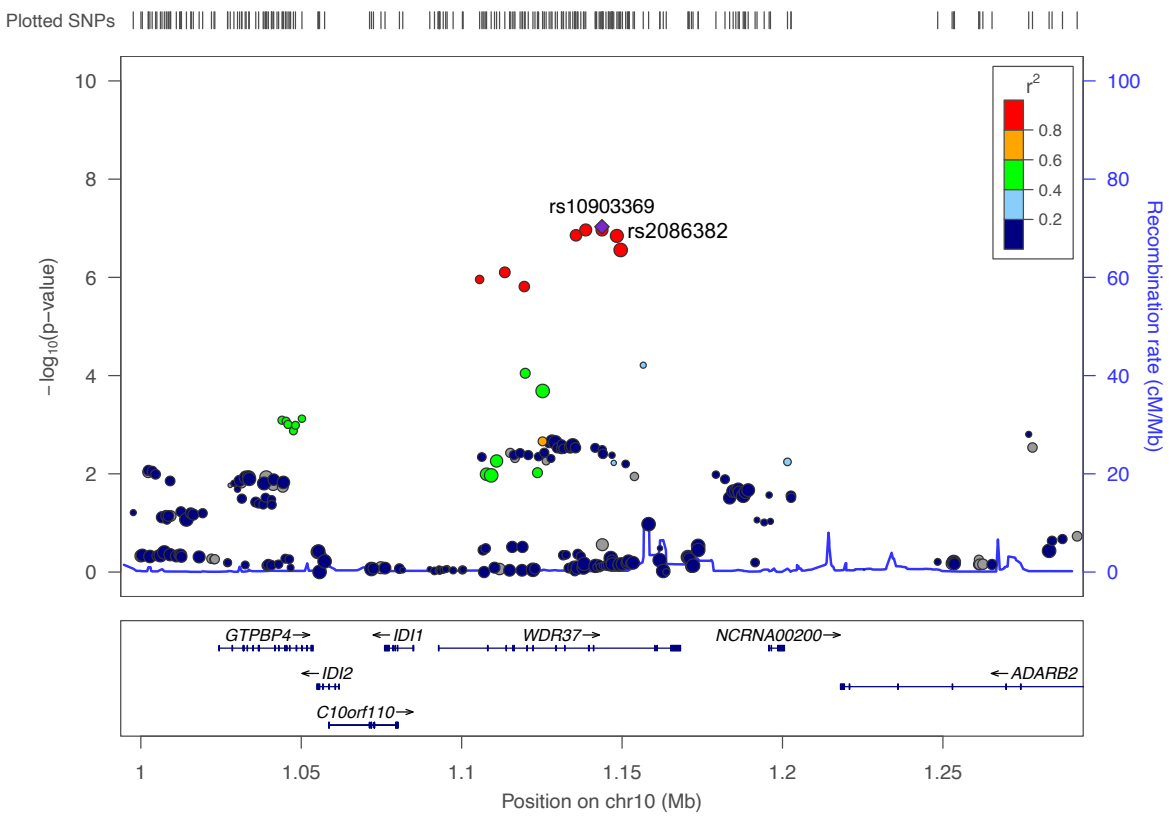

Figure 4.5. Regional association plot showing SNPs (MAF>0.05) in LD with rs10903369

Plot is based on univariate GWAS results. rs2086382 (lead SNP in multivariate analysis) is also shown. SNPs with missing LD information are shown in grey. Mb=megabase. r^2=LD between SNPs. Circle sizes reflect sample size for each SNP.

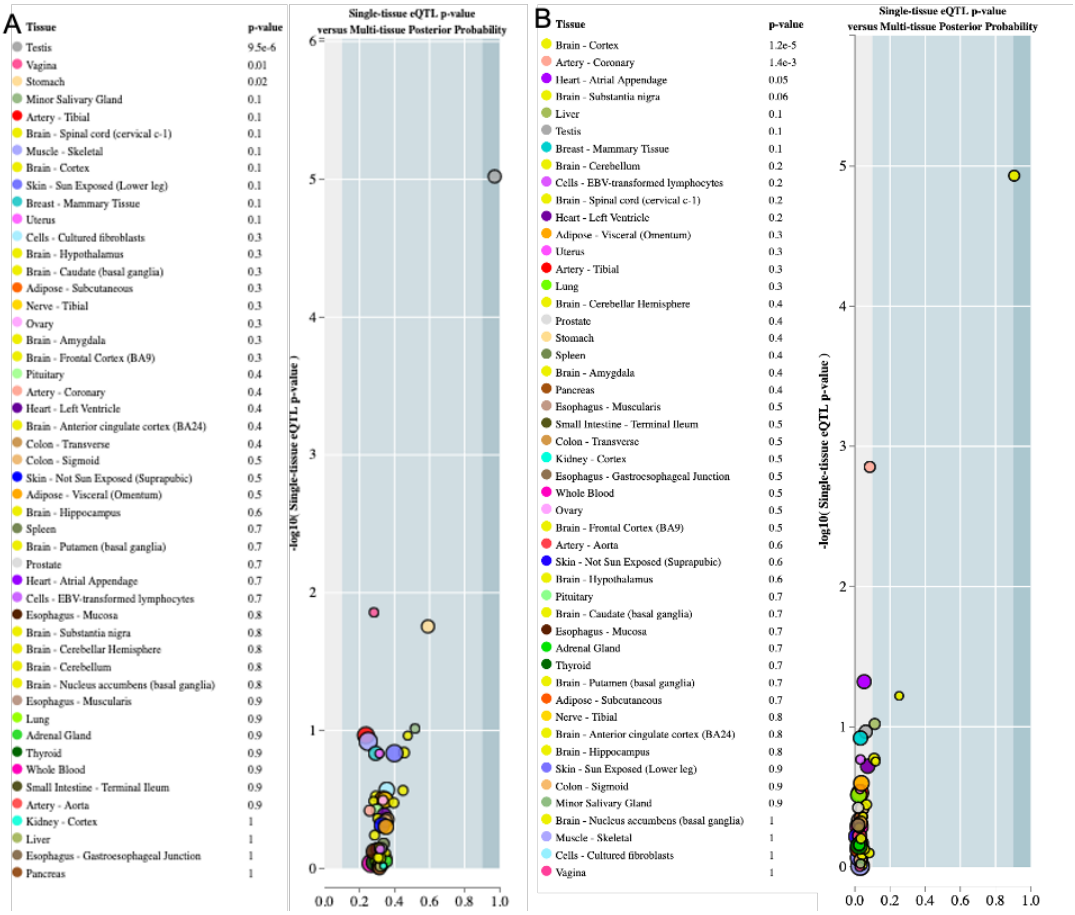

Figure 4.6. Multi tissue eQTLs for rs10903369 in *WDR37* (A) and *IDI1* (B) *P*-value=single tissue eQTL *P.* 0-1=normalised effect sizes (NES).

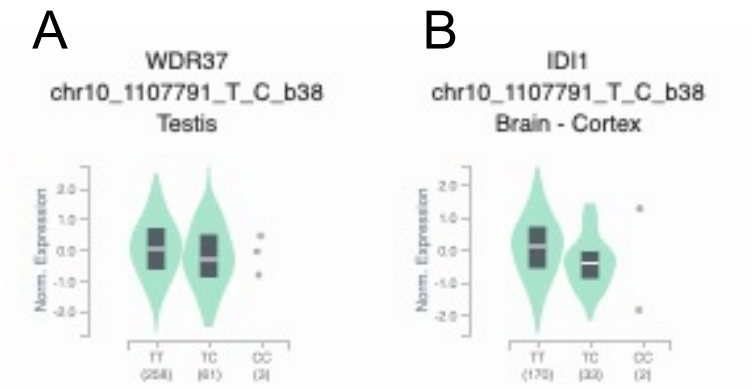

Figure 4.7. Violin plots for eQTL regulation of (A) *WDR37* and (B) *IDI1* by rs10903369

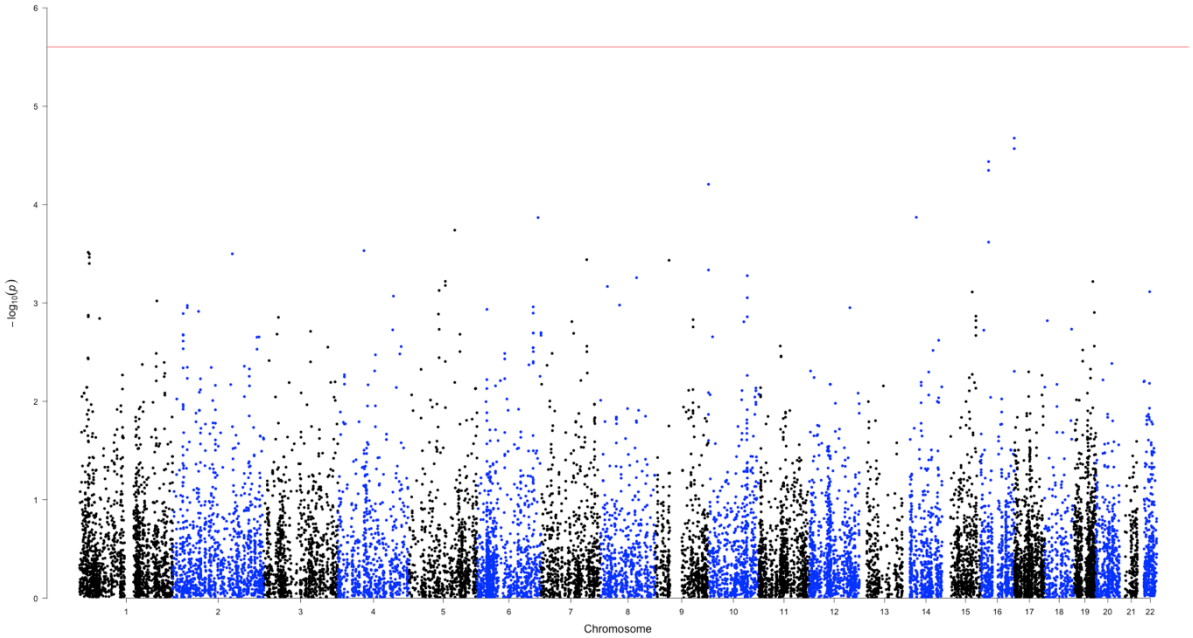

Figure 4.8. Gene analysis from univariate GWAS outputs (n=1,649)
Statistical significance threshold is P=2.5 x 10^{-6} due to correction for multiple testing using the Bonferroni correction for 20,000 genes (red line).

Table 4.9. Gene-analysis for response to oxaliplatin-based chemotherapy

Gene ID	Gene Symbol	Gene Name	Chr	SNPs (n)	*P*	*Q*
Univariate						
2622	*GAS8*	Growth arrest specific 8	16	24	$1.6x10^{-5}$	0.09
22884	*WDR37*	WD repeat domain 37	10	150	$2.2x10^{-5}$	0.09
79007	*DBNDD1*	Dysbindin domain containing 1	16	23	$2.3x10^{-5}$	0.09
10261	*IGSF6*	Immunoglobulin superfamily member 6	16	16	$2.7x10^{-5}$	0.09
51108	*METTL9*	Methyltransferase like 9	16	25	$3.2x10^{-5}$	0.09
89874	*SLC25A21*	Solute carrier family 25 member 21	14	370	$1.4x10^{-4}$	0.31
135138	*PACRG*	Parkin coregulated	6	804	$1.6x10^{-4}$	0.31
146183	*OTOA*	Otoancorin	16	28	$1.9x10^{-4}$	0.31
100505841	*LOC100505841*	Zinc finger protein 474-like	5	14	$2.0x10^{-4}$	0.31
7364	*UGT2B7*	UDP glucuronosyltransferase family 2 member B7	4	130	$2.3x10^{-4}$	0.32
Multivariate						
2622	*GAS8*	Growth arrest specific 8	16	24	$2.6x10^{-5}$	0.21
10184	*LHFPL2*	LHFPL tetraspan subfamily member 2	5	116	$3.9x10^{-5}$	0.21
79007	*DBNDD1*	Dysbindin domain containing 1	16	23	$4.6x10^{-5}$	0.21
115352	*FCRL3*	Fc receptor like 3	1	62	$1.9x10^{-4}$	0.47
115416	*MALSU1*	mitochondrial assembly of ribosomal large subunit 1	7	4	$2.4x10^{-4}$	0.47
4012	*LNPEP*	Leucyl and cystinyl aminopeptidase	5	281	$2.8x10^{-4}$	0.47
51752	*ERAP1*	Endoplasmic reticulum aminopeptidase 1	5	79	$2.9x10^{-4}$	0.47

3950	*LECT2*	Leukocyte cell derived chemotaxin 2	5	188	$3.2x10^{-4}$	0.47
54212	*SNTG1*	Syntrophin gamma 1	8	980	$3.9x10^{-4}$	0.47
72	*ACTG2*	Actin gamma 2, smooth muscle	2	88	$4.1x10^{-4}$	0.47

Statistical significance was corrected for multiple testing using the Bonferroni correction for 20,000 genes to P=2.5 x 10^{-6}. Q=Q-value to take into account false discovery rate. Ten genes with the smallest P-values for each GWAS shown.

Table 4.10. Gene-set analysis for response to oxaliplatin-based chemotherapy

	GO ID	GO term	Genes (n)	*Q*
Univariate				
	GO:0003887	DNA directed DNA polymerase activity	23	0.82
	GO:0034061	DNA polymerase activity	28	0.99
	GO:0003279	Cardiac Septum Development	103	0.99
	GO:0005112	Notch binding	19	0.99
	GO:0090335	Regulation of brown fat cell differentiation	14	0.99
Multivariate				
	GO:0044255	Cellular lipid metabolic process	717	0.33
	GO:0006629	Lipid metabolic process	844	0.39
	GO:0016758	Transferase activity, transferring hexosyl groups	151	0.39
	GO:0019585	Membrane protein intracellular domain proteolysis	19	0.55
	GO:0016071	mRNA metabolic process	15	0.55

Competitive gene-set analysis performed with all gene results from univariate and multivariate gene analysis. Statistical Significance Q=0.05. Q=Q-value to take into account false discovery rate. Five gene-sets with the smallest Q-values for analysis shown.

4.4 Discussion

4.4.1 Discovery of novel variants suggestive of association with response to chemotherapy

In this investigation, one key locus located at 10p15.3 (lead SNP rs10903369) is suggestive of association with response to chemotherapy with supporting evidence in relation to a significant effect on OS. The minor allele (C) frequency for this SNP in COIN and COIN-B was 9%, compared to the Allele Frequency Aggregator (ALFA) European MAF of 8% (Phan et al., 2020). Individuals with both copes of the minor allele are approximately two fold more likely to respond to treatment (93% compared to 55% for homozygous major allele) and have a median survival time 3.8 months longer than individuals who are homozygous for the major allele. There are no previously reported associations with any phenotype, including cancer outcomes, for this SNP (Sherry et al., 2001).

4.4.1.1 The effect of additional covariates

With the addition of clinical covariates that may affect CRC outcomes, there is some change in the effect size of associations with response to oxaliplatin-based chemotherapy compared to associations of rs10903369 and response alone. It is likely that the tumour biomarker mutation status, which were shown to be significantly associated with response in this cohort (Table 4.1), had a confounding effect on response to treatment.

KRAS mutations have been established to have an effect on anti-EFGR treatments such as cetuximab but this study demonstrates that they may also be playing a role in chemotherapy efficacy (Karapetis et al., 2008). The evidence surrounding the effect of *KRAS* mutations on oxaliplatin without cetuximab or panitumumab are conflicting. There is some evidence that *KRAS* mutant patients have worse outcomes which indicate poor response to chemotherapy (Zocche et al., 2015, Smith et al., 2013c). However, other investigations indicate the opposite or no predictive role these of these mutations (Sharma et al., 2010, Bokemeyer et al., 2009, Lin et al., 2014).

There have been reports of patients with *BRAF* mutant CRCs having low response to oxaliplatin and survival (OS and PFS) rates regardless of disease stage (Clarke and Kopetz, 2015, Souglakos et al., 2009). This supports the findings of this investigation as shorter PFS survival indicates a poor response to treatment, as RECIST 1.0 guidelines describe non-responders as having stable or progressive disease (McShane et al., 2005).

There are fewer studies into the effect of *NRAS* on any CRC phenotype, with limited evidence in smaller cohorts that *NRAS* mutant tumours could predict worse prognosis in CRC patients (Hu et al., 2018). This, coupled with the moderate *P*-values for associations between *NRAS* mutation status and response (Table 4.1), should be considered before drawing any conclusions in relation to the role of *NRAS* in response to oxaliplatin-based chemotherapy.

Whilst it has been established that MSI can be associated with resistance to 5-FU, a key component of FOLFOX, it was not possible to investigate this within the sample due to limited success in the identification of MSI, with only 66 patients from COIN and none from COIN-B (where only FOLFOX was utilised) identified with MSI (Popat et al., 2005, Jo and Carethers, 2006).

It is, however, important to note the difference between the effect size (OR for response and HR for survival) when covariates such as these *RAS* and *RAF* mutations on rs10903369, are taken into account, is not statistically significant. This indicates that whilst these mutations may have an effect on response independently, they are having limited impact on the reported associations in this chapter.

4.4.1.2 Different treatment types

Whilst there is a difference in the reported associations between rs10903369 and response to treatment depending on type of chemotherapy and whether patients are given cetuximab, there is not a significant difference between any of these subgroups. This may be due to similar effect sizes in the same direction and a larger standard error than the whole cohort, a potentially due to the decreased sample size (Altman

and Bland, 2005). Equally, as capecitabine is designed to mimic the continual infusion of 5-FU, it could be argued that this results in both treatments having similar mechanisms of action and therefore no difference in effect based on chemotherapy type would be expected (Twelves et al., 2001).

Additionally, there is no significant difference on the effect of this SNP on response and OS between patients who were given cetuximab and those who were given chemotherapy alone. This is supported by the heterogeneity of rates of response in this cohort (Chapter 2, Sections 3.3 and 4.1).

4.4.1.3 *WDR37* and *IDI1*

Reduced expression of *WDR37* has been linked with the reported variation in rs10903369 (The GTEx Consortium, 2013). *WDR37* produces a protein of unknown function which has been linked to developmental disorders that are known to affect brain and facial development (Reis et al., 2019). There have been no links demonstrated between the transcribed protein and any CRC phenotype.

A circular RNA molecule, hsa_circ_0004277, found within an intron of *WDR37*– which is in eQTL with rs10903369 – has been shown to contribute to malignancy phenotypes in CRC(Yang et al., 2020). This contribution is through ‘microRNA sponging' - the competitive inhibition of microRNA by circular RNA molecules resulting in microRNA knockdown (Qu et al., 2015, Hammond, 2007). Yang et al. (2020) demonstrated that hsa_circ_0004277 sponges miR-512-5p resulting in the upregulation of *Human prothymosin-α (PTMA),* however this does require further biological validation. If this mechanism can be confirmed, reduced expression of hsa_circ_0004277, may result in downregulation of *PTMA*.

PTMA plays a key role in tumorigenesis, with evidence that it influences the expression of the known tumour suppressor, *TP53* (Zhang et al., 2014b). Decreased *PTMA* expression results in malignant phenotypes including reduced rates of apoptosis and increased cell proliferation, which could contribute to a no response phenotype (Jiang et al., 2003).

Whilst this circular RNA is within a gene which has been is in eQTL with rs10903369, there is no data on if hsa_circ_0004277 itself is in eQTL with this SNP. Therefore experimental confirmation of the potential relationship between variation in rs10903369, hsa_circ_0004277 and the response phenotypes are required. Futthermore, the region in which this circular RNA is found is ~27kB away from the rs10903369. Additionallly, the SNPs in the circular RNA are monomorphic within patients from COIN and COIN-B. Therefore, hsa_circ_0004277 is unlikely to be contributing to the effect from this locus.

Despite the limited evidence to support hsa_circ_0004277's relationship with rs10903369 and response to oxaliplatin-based chemotherapies, the relative proximity of the circular RNA to the SNP is important to note. Therefore this locus should not be dismissed as a potential signpost for further investigation.

IDI1 is downregulated in the presence of the minor allele of rs10903369. *IDI1* catalyses the of isomerisation of isopentenyl disphosphate (IPP) to dimethylallyl diphosphate (DMAPP) as part of isoprenoid biosynthesis. This may play a role in resistance to oxaliplatin which may result in a no response phenotype in CRC cell lines through its contribution to the mevalonate pathway (Cornforth et al., 1966, Caruso and Notarnicola, 2005). The mevalonate pathway has been shown to have increased activity in CRC tissue compared to wild-type cells (Virag et al., 2013).

Whilst there have been no published associations between variants in this gene and any CRC outcome, these biological findings could indicate that this could be a promising biomarker for response to oxaliplatin-based chemotherapy. Especially when considering that this resistance pathway may be downregulated, supporting the increased response to oxaliplatin in the presence of variation at rs10903369.

This work has identified mechanisms that may link both *IDI1* and *WDR37* to the observed response phenotypes. However, it is important to note that the that tissues significantly associated with differential regulation in tissue types that would not appear to be relevant to CRC (brain and testes, Figure 4.7). The evidence to support these associations influencing response phenotypes is therefore limited. However, as GTEx only uses expression data from healthy individuals, the lack of association in

tissues such colon or rectum may not necessarily mean that the regulation of these genes is not affected by this SNP within CRC tumours. It has been observed that cancerous tissues can have different expression profiles to their corresponding healthy tissue type (The GTEx Consortium, 2013, Yang et al., 2017b, Kheirelseid et al., 2013). Because of this, the true relevance of these observed changes in expression of *IDI1* and *WDR37* to response to oxaliplatin-based chemotherapy is unknown until exploration of the influence of rs10903369 on these genes in CRC tissues can be undertaken.

4.4.2 Gene-based and gene-set analyses

No genes or gene-sets are statistically associated with response either in a univariate and multivariate model after correction for multiple testing. However, it has been argued that the Bonferroni correction for multiple testing can be overly conservative in large data sets such as this (due to genes and gene-sets not being independent) and therefore actual associations could have been missed (Bender and Lange, 1999, Bland and Altman, 1995).

In an attempt to account for the conservative nature of the Bonferroni correction and discover if any of the biologically significant results could be statically significant, *Q*-scores, which estimate a more generous but still conservative FDR, were also used (Wright, 1992). There are still no genes or gene-sets that reach the FDR-adjusted significance threshold. This may strengthen the argument that there are no true statistical associations with genes or gene-sets for response to oxaliplatin within this cohort.

It is important, however, to note that some genes with the lowest *P*-values are members of family that are associated with CRC development of progression. Genes within the *Growth-arrest specific (GAS)* family including *GAS8*, which have the lowest *P*-value in both the univariate and multivariate gene analyses, have been suggested to play a role within the development of gastric cancers (Esfandi et al., 2019). Interestingly, downregulation of *GAS1*, another member of this family has been shown to correlate with recurrence in patients with stage II and II CRC (Jiang et al., 2011).

WDR37 is also found within the gene analyses, which is likely due to the loci suggestive of significance and other variants nearby which are not in linkage with this particular loci.

4.4.3 Conclusion and follow up studies

This investigation has resulted in the identification of a novel locus suggestive of significant association with both response to oxaliplatin-based chemotherapy and OS in a region where is there is some evidence of activity in relation to colorectal cancer. This, therefore provides a potential signpost for further investigation.

Whilst this result is promising, this study shows few germline variants that are suggestive of association with response to oxaliplatin-based chemotherapy. Modest sample sizes, like that of this investigation (n=1,649) have been shown to limit the power of GWASs (Hong and Park, 2012). However, the lack of significant data may be due to biological response of the tumours. It may also be the case that tumour genotypes have greater influence on response to oxaliplatin-based chemotherapy, as large associations between *BRAF*, *KRAS* and *NRAS* mutational status have been demonstrated in this investigation and in other published studies (Souglakos et al., 2009, Clarke and Kopetz, 2015).

In order to strengthen the conclusions of this investigation, single SNP validation of the associations between rs10903369 and response to chemotherapy would have to be performed in an independent cohort. In order to detect this association (Univariate OR=2.10, Multivariate OR=2.11) at a significance level of $P \leq 0.05$ with a power of 80%, a minimum sample size of 373 patients would be require. This involves the assumption of similar rates response (~60%) and MAF of 0.1.

Whilst oxaliplatin-based chemotherapy is an important treatment for advanced colorectal cancer, it is are not the only treatment currently available. Therefore it will be important to investigate possible germline associations with other approved

treatments, including anti-EFGR monoclonal antibody cetuximab which will be explored later in this thesis (Chapter 5).

5. Chapter 5: Cetuximab improves response in patients with RAS wild type advanced colorectal cancer and identification of potentially predictive biomarkers

5.1 Introduction

Epidermal growth factor receptor (EGFR), is a transmembrane protein and initiator of the EGFR pathway (Chapter 1, 1.3.2.1.4). Activation of this pathway, through ligands binding to EGFR or mutations in genes downstream of the receptor, result in dysregulated growth, differentiation and survival of cells – a phenotype often observed in multiple epithelial cancers (Oda et al., 2005, Cohen et al., 1980, Sigismund et al., 2018). Genes downstream of the receptor (including *KRAS*, *NRAS* and *BRAF*) are often activated or upregulated in cases of more aggressive CRC where patients have worse response to treatments and poor prognosis (Benvenuti et al., 2007, Eklöf et al., 2013).

Monoclonal antibodies against EGFR have been shown to be beneficial in the treatment of *KRAS* or *RAS* [*KRAS* and *NRAS*] wild-type CRCs both as a monotherapy or when used in combination with chemotherapy (Guren et al., 2017, Karapetis et al., 2008, Li et al., 2020, Stintzing et al., 2016, Khattak et al., 2015). In a recent meta-analysis of multiple randomly controlled trials (RCTs), cetuximab was shown to significantly improve OS (HR=0.74, 95% CI=0.55–0.98, P=0.04), PFS (HR=0.63, 95% CI=0.50–0.79, P<0.01), overall recurrence rate (ORR=1.76, 95% CI=1.40–2.21, P<0.01) and resection rate (RR=2.03, 95% CI=1.25–3.29, P<0.01) in patients with *KRAS* wild-type tumours with liver metastasis (Lv et al., 2017).

5.1.1 Cetuximab benefit in the COIN and COIN-B trials

Unexpectedly, the COIN trial of oxaliplatin and fluoropyrimidine chemotherapy with or without cetuximab in the first-line treatment of aCRC, found no evidence for cetuximab benefit in patients with *KRAS* wild-type CRCs, in terms of both PFS and OS (Chapter

1, Section 1) (Maughan et al., 2011). Exploratory analyses suggested a potential interaction (*P*=0.10) with the type of co-administered chemotherapy; with fluorouracil-based chemotherapy (FOLFOX) favouring cetuximab (PFS HR=0.72, 95% CI=0.53–0.98, *P* not reported) and capecitabine-based chemotherapy favouring no cetuximab (HR=1.02, 95% CI=0.82–1.26, *P* not reported). Addition of cetuximab increased toxic side effects including diarrhoea, skin rash, lethargy and peripheral neuropathy (all grades of toxicity, Fisher's exact test *P*<0.05) (Adams et al., 2009). Based on this, it is important to determine which patients benefit from cetuximab to reduce unnecessary toxic side effects that can result in treatment delay or cessation.

5.1.2 Clinical guidelines for cetuximab in colorectal cancer

Based on the evidence of lack of benefit in patients with *RAS* tumours, cetuximab is indicated for the treatment of patients with colorectal cancers that do not harbour somatic mutations in exon 2 (codons 12 and 13), exon 3 (codons 59 and 61), and exon 4 (codons 117 and 146) of *KRAS* or *NRAS* [*RAS*] (Eli Lilly and Co, 2021). Current NICE guidelines for cetuximab in colorectal cancer are similar – cetuximab is recommended for previously untreated *RAS* wild-type, EGFR-expressing mCRCs in combination with either FOLFOX or FOLFIRI (National Institute for Health and Care Excellence, 2017). The NICE guidance for cetuximab was due to be reviewed in 2020, a time that coincided with the modification of cancer treatment due to the Covid-19 pandemic (National Institute for Health and Care Excellence, 2020b).

5.1.3 Hypothesis and aims

The main hypothesis of this chapter is that within a subgroup of patients who have better response to cetuximab, there are germline variants associated with this response phenotype. The specific aims are:

- Perform an exploratory analyses of clinicopathological factors to determine a subset of patients with increased response to cetuximab.
- Search for germline SNPs associated with response to cetuximab in more responsive patients.
    - Perform a univariate GWAS for response to cetuximab on a subgroup of genotyped patients from COIN and COIN-B.
    - Perform a multivariate GWAS for response to cetuximab on the same patients to include known prognostic factors.
- Perform survival and interaction analyses on SNPs suggestive of association with response to cetuximab.
- Determine if there is any supporting evidence for newly discovered SNPs.

5.2 Methods

5.2.1 Patient inclusion and endpoints

Of the 2,671 patients recruited into COIN and COIN-B (Chapter 1, Sections 1.4.4 and 1.4.5), 1,950 patients had complete germline genetic information. Of these, 1,948 had survival data and 1,649 response data (at 12 weeks; Chapter 4, Section 2.1).

5.2.1.1 Response at 12 weeks

Assessment of response was performed at 12 weeks, since at this point patients from all trial arms received identical levels of chemotherapy with or without cetuximab. This was also prior to any interruption to treatment for the intermittent therapy arms (C, D and E). Response was defined as complete or partial response using RECIST 1.0 guidelines and no response was defined as stable or progressive disease (Therasse et al., 2000).

5.2.1.2 Overall survival

I initially considered OS as an alternative measure of cetuximab efficacy. Patients with *RAS* wild-type CRCs who received continuous chemotherapy alone had better OS (median 20.2 months) compared to patients who received only intermittent chemotherapy (median 16.0 months, HR=0.80, 95% CI=0.66–0.97, *P*=0.03, Figure 5.1). No difference was observed in patients with *RAS* mutant CRCs (continuous chemotherapy alone, median 16.5 months; intermittent chemotherapy alone, median 16.7 months; HR=0.95, 95% CI=0.77–1.19, *P*=0.68, Figure 5.1).

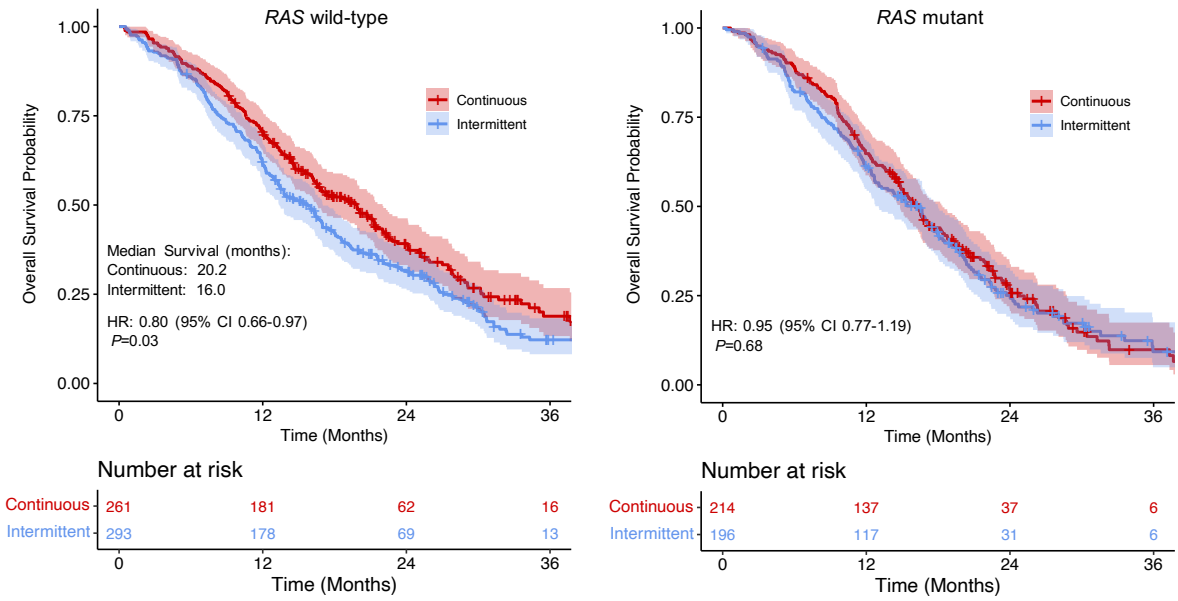

Figure 5.1. Kaplan-Meier curves showing benefit of continuous versus intermittent chemotherapy on OS by *RAS* mutation status (excludes patients with cetuximab)

(A) Patients with *RAS* wild-type CRCs (n=554), and, (B) patients with *RAS* mutant CRCs (n=410). *P*-values show difference between treatment groups, with shaded areas representing 95% CIs.

A potential predictive effect of cetuximab based on these scheduling backgrounds was also explored. No clear benefit for cetuximab was seen in patients with *RAS* wild-type CRCs who received continuous therapy (OS median 20.1 months, HR=0.94, 95% CI=0.76–1.16, *P*=0.6), although in patients with intermittent therapy, a trend towards cetuximab benefit was observed (median 20.3 months, HR=0.80, 95% CI=0.61–1.04, *P*=0.09, Figure 5.2).

Importantly, there was a clear benefit for continuous treatment in patients with *RAS* wild-type CRCs who did not receive cetuximab (median OS increase >4 months) as compared to patients who received intermittent treatment. As only 47% of patients with chemotherapy alone but 71% with chemotherapy plus cetuximab had continuous treatment, this benefit would favour better OS in the cetuximab group (Table 5.1). Due to this bias, OS could not be used as an accurate marker of cetuximab efficacy moving forwards. This data has been presented in this section so as not to detract from the results section of this chapter.

5.2.2 Exploratory analysis of clinicopathological factors

Logistic regression analyses for response to treatment and ten clinicopathological factors (sex, age, site of primary tumour, number of metastatic sites, presence of only liver metastases, synchronous metastases, chemotherapy type, white blood cell (WBC) count, *RAS* mutation status and *BRAF* mutation status) was performed in STATA/SE version 16.1. An interaction effect between each factor and cetuximab was determined using a nominal significance threshold of $P_{Interaction}<0.05$. A Forest plot for associations was visualised using the *ipdmetan* package in STATA (Fisher, 2014).

5.2.3 Survival analyses

Survival outcomes were analysed in R-version 3.2.5 (R Core Team, 2018). For validation of SNPs suggestive of association with response at 12 weeks, OS was analysed by univariate and multivariate Cox proportional hazards models using the *survival* package (Therneau, 2020, R Core Team, 2018). Survival curves were plotted

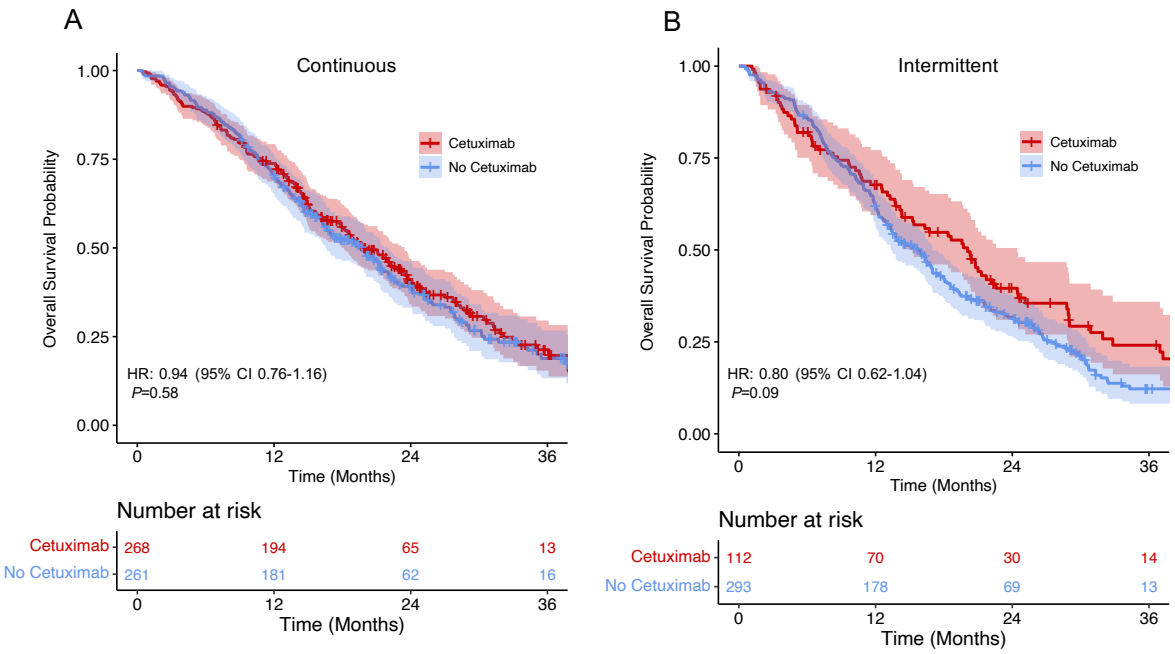

Figure 5.2. Kaplan-Meier curves showing benefit of cetuximab on OS in *RAS* wild-type patients by treatment scheduling

(A) Continuous chemotherapy (n=529), and, (B) intermittent chemotherapy (n=405). *P*-values show difference between treatment groups, with shaded areas representing 95% CIs.

Table 5.1. Treatment scheduling in patients with *RAS* wild- type CRCs

	Cetuximab	No Cetuximab
Continuous chemotherapy	268 (71%)	261 (47%)
Intermittent chemotherapy	112 (29%)	293 (53%)
All	380 (100%)	554 (100%)

Percentages shown in parentheses.

using the Kaplan-Meier method using the *survminer* and *ggplot2* packages in R (Therneau, 2020, Wickham, 2016).

5.2.4 Genome Wide Association Studies

Details of methodology for univariate and multivariate additive logistic GWAS using PLINK version 2.0 can be found in Chapter 2, Section 4.3 (Purcell et al., 2007). Patients who did not receive cetuximab and those who received cetuximab but had *RAS* mutant tumours were excluded from this study. SNPs MAF<5% were also excluded from analyses. The interactive effect ($P_{Interaction}$) for SNPs was calculated using data from all *RAS* wild-type patients with complete response data (n=794, Figure 5.3) with cetuximab status as a factor using the --logistic interaction command in PLINK. For the multivariate analysis, all covariates were included in the interactive model (Table 5.2).

Additive logistic regressions to test for association between known covariates and response to cetuximab (Table 5.2) were carried out using the -no-snp function. These covariates were also included in the multivariate GWAS.

ORs reported in Table 5.3 are based on two-sided suggestive ($1x10^{-5}$) and genome wide ($5.0x10^{-8}$) significance thresholds (α), a power (1-β) of 80%, response rate of 71% (response rate within this subgroup) and MAF >20%. These were calculated using the *genpwr* package in R-version 3.5.2 and described further in Chapter 2, Section 4.4 (Moore and Jacobson, 2020, Hong and Park, 2012, R Core Team, 2018).

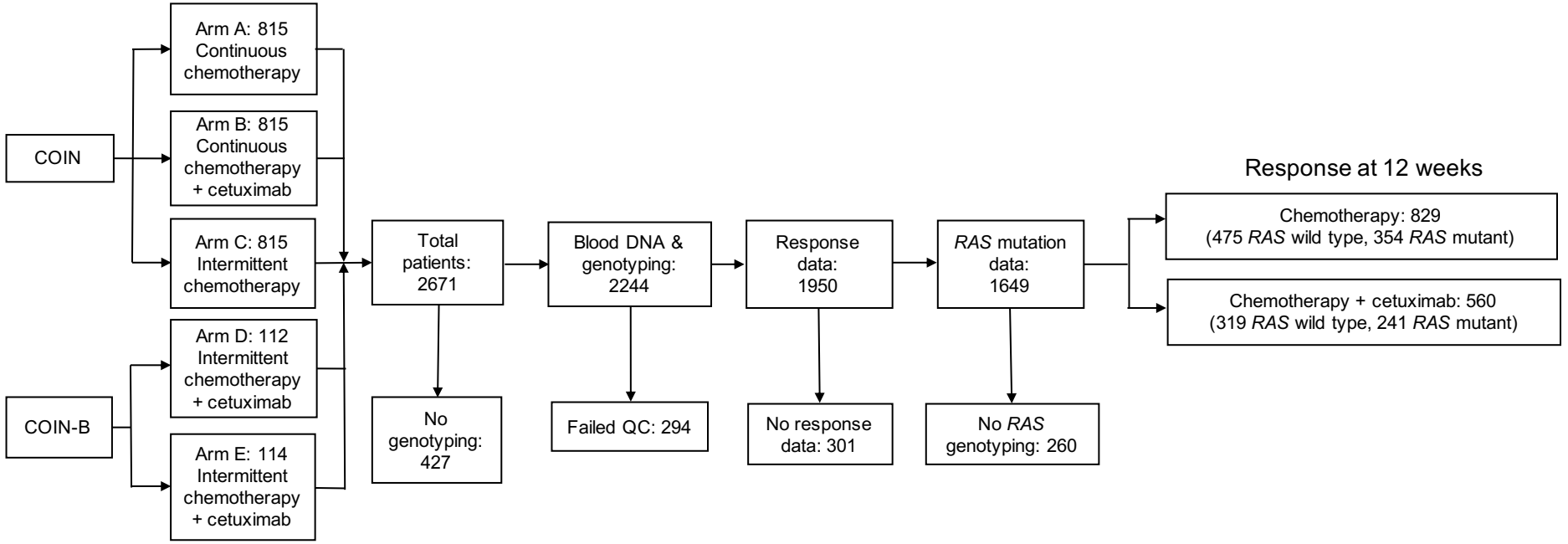

Figure 5.3. CONSORT diagram of the analysis strategy

COIN patients were randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy (Arm A, n=815), continuous chemotherapy with cetuximab (Arm B, n=815), or intermittent chemotherapy (Arm C, n=815). COIN-B patients were randomised 1:1 to receive intermittent chemotherapy and cetuximab (Arm D, n=112) or intermittent chemotherapy and continuous cetuximab (Arm E, n=114). Of these, 2,244 were genotyped on Axiom arrays, 1,950 passed QC, 1,649 had response data and 1,389 had *RAS* genotyping data. I considered response at 12 weeks which was prior to any interruption to treatment for the intermittent therapy arms – in total, 829 patients had chemotherapy alone (475 were *RAS* [*KRAS* and *NRAS*] wild-type and 354 *RAS* mutant) and 560 had chemotherapy plus cetuximab (319 were *RAS* wild-type and 241 *RAS* mutant).

Table 5.2. Variables included in the multivariate analyses of response to cetuximab and overall survival

	Response (n=309) OR (95% CI)	***P***	**OS (n=367) HR (95% CI)**	***P***
Age	1.00 (0.97–1.03)	0.93	1.00 (0.99–1.02)	0.73
Sex				
Male	1.00		1.00	
Female	0.73 (0.42–1.26)	0.25	0.92 (0.70–1.21)	0.56
Disease site				
Colon	1.00		1.00	
Rectum	1.25 (0.70–2.24)	0.45	1.21 (0.92–1.59)	0.18
Number of metastatic sites				
0-1	1.00		1.00	
≥2	0.96 (0.72–1.28)	0.79	1.63 (1.24–2.12)	$3.7x10^{-4}$
WHO performance status				
0-1	1.00		1.00	
2	1.27 (0.36–4.57)	0.71	1.83 (1.04–3.20)	0.04
Primary tumour resected				
No	1.00		1.00	
Yes	0.91 (0.52–1.59)	0.74	0.80 (0.60–1.05)	0.11
Local recurrence	0.79 (0.25–2.51)	0.69	0.70 (0.40–1.26)	0.24
White cell count	0.93 (0.86–1.01)	0.09	1.07 (1.04–1.11)	$5.3x10^{-6}$
BRAF mutation status				
Wild-type	1.00		1.00	
Mutant	0.32 (0.15–0.65)	$1.6x10^{-3}$	3.10 (1.50–6.41)	$2.2x10^{-3}$
Chemotherapy regimen				
XELOX	1.00		1.00	
FOLFOX	1.15 (0.68–1.94)	0.60	0.90 (0.70–1.17)	0.44

Multivariate logistic regression analyses (for response: responders – complete or partial response vs non-responders – stable or progressive diseases) and multivariate Cox regression (for OS) of prognostic factors were performed. Analysis excludes cases with missing values for any prognostic factor or response data at 12 weeks. Odds and hazard ratios for each variable were adjusted for all other covariates listed. Positive ORs indicate an increased likelihood of response and positive HRs indicate an increased likelihood of early death. Odds and Hazard Ratios for reference groups are listed as 1.0.

Table 5.3. Detectable Odds Ratios at 80% power in logistic regression analysis of response at 12 weeks

Test	No. of cases	No. of responders	*P*	Detectable OR
Univariate additive GWAS				
	319	228	$1x10^{-5}$	4.1
	319	228	$5x10^{-8}$	6.2
Multivariate additive GWAS				
	309	223	$1x10^{-5}$	4.3
	309	223	$5x10^{-8}$	6.5

Detectable odds ratios (ORs) for each SNP take into account the minimum MAF (20%), sample size, probability of failure and adjusted two-sided (α) suggestive significance ($P<1.0x10^{-5}$) and genome wide significance ($P<5.0x10^{-8}$). All calculations were based on a power (1-β) of 80%.

5.2.5 Gene-based and gene-set analyses

MAGMA version 1.8, NCBI 37.3 gene definitions and predefined gene-sets (~8,000) were used to perform gene and gene-set analyses (de Leeuw et al., 2015). Gene analyses were run under the snpwise model and gene-set analysis under the competitive model. The Bonferroni adjusted significance threshold of $P<2.5x10^{-6}$ (correction for 20,000 genes) was used for gene analyses. Both gene and gene-set analyses were also analysed for significance using a *Q*-value which adjusted for FDR with a standard significance threshold of $Q<0.05$.

5.2.6 Downstream bioinformatic analyses

Regional association analyses of SNPs were performed using LocusZoom (Chapter 2, Section 2). Identification of eQTLs was performed using the GTEx project database (Chapter 2, Section 2).

5.3 Results

5.3.1 Exploratory analysis of clinicopathological factors

An exploratory analyses of clinicopathological factors (n=10 factors) was performed on all patients with complete genotypes and data for response at 12 weeks (n=1,649 patients) from COIN and COIN-B to identify predictive biomarkers for response to cetuximab (Figure 5.3). 994 patients received chemotherapy alone (475 had *RAS* wild-type CRCs, 354 *RAS* mutant CRCs and 165 were not genotyped) and 655 received chemotherapy with cetuximab (319 had *RAS* wild-type CRCs, 241 had *RAS* mutant CRCs and 95 were not genotyped).

RAS mutation status was predictive for response to cetuximab ($P_{Interaction}$<0.01, Figure 5.4). In patients with *RAS* wild-type CRCs, cetuximab improved response with 72% (228/319) responding compared to 61% (289/475) with chemotherapy alone (OR=1.61, 95%CI=1.19–2.19, *P*<0.01). A detrimental effect for cetuximab was seen in patients with *RAS* mutant CRCs with 46% (112/241) of these patients responding to chemotherapy with cetuximab as compared to 54% (192/354) without cetuximab (OR=0.72, 95% CI=0.52–1.00, *P*=0.05).

Type of co-administered chemotherapy was not significantly associated with response to cetuximab in a subset of patients with *RAS* wild-type CRCs ($P_{Interaction}$=0.82). Seventy-one percent (108/152) of patients treated with XELOX plus cetuximab responded versus 60% (184/307) with XELOX alone (OR=1.64, 95% CI=1.08–2.49, *P*=0.02) and 81% (120/149) of patients treated with FOLFOX plus cetuximab responded versus 63% (105/168) without cetuximab (OR=1.53, 95%CI=0.97–2.43, *P*=0.07).

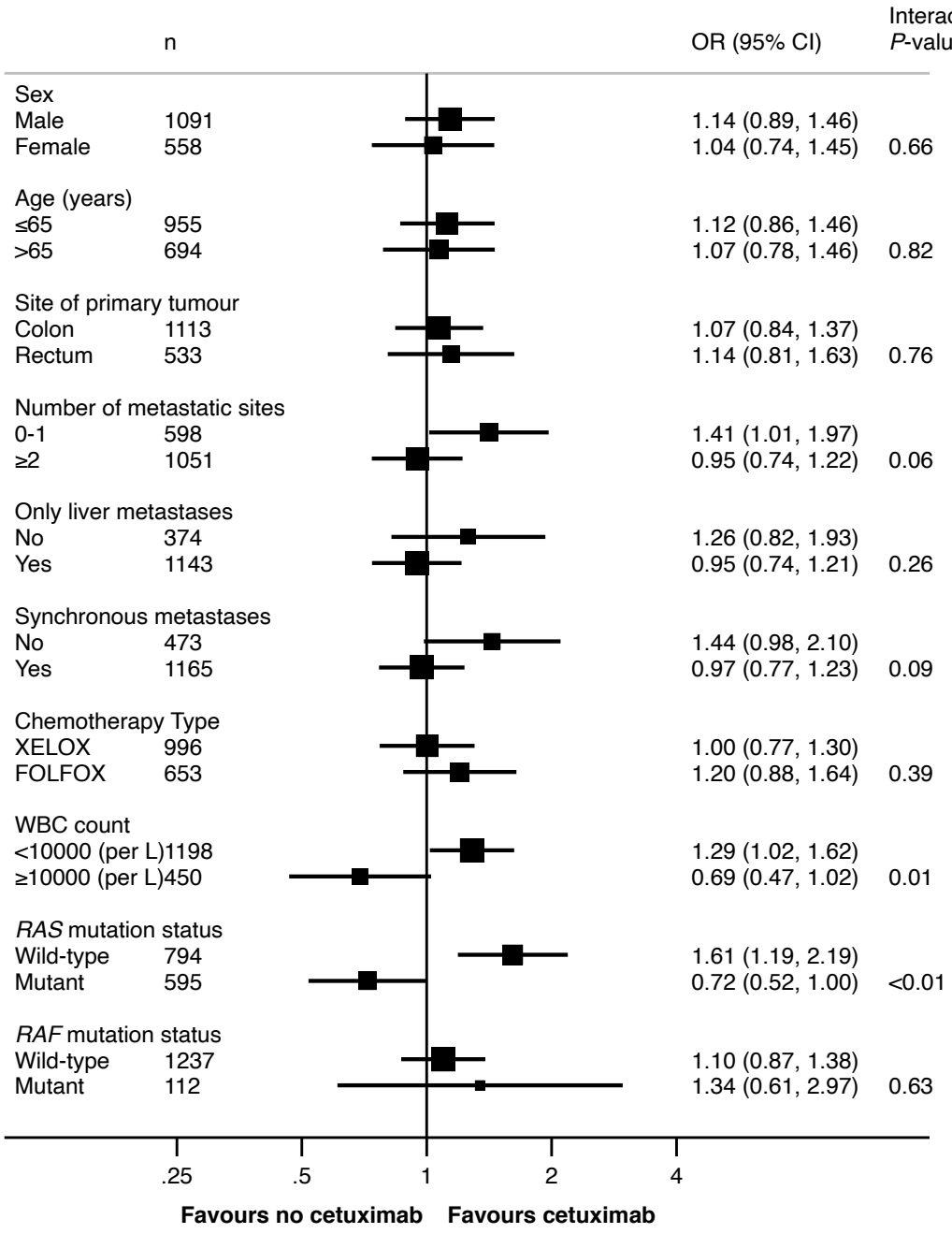

Figure 5.4. Exploratory analysis of clinicopathological factors for response to cetuximab

Forest plot for response at 12 weeks (patients with complete or partial response were compared to those with stable or progressive disease as defined by REICIST 1.0 guidelines, n=1,649). Interactive *P*-values were calculated for the interaction between cetuximab and each factor. HR=Hazard Ratio, WBC=White blood cell count.

Raised WBC (≥10,000 per L) also predicted worse response to cetuximab at 12 weeks. In patients with raised WBC (≥10,000 per L), 51% (78/152) of patients treated with chemotherapy plus cetuximab responded to treatment, compared to 60% (180/298) of patients treated with chemotherapy alone (OR=0.69, 95% CI=0.47–1.02, P=0.07). In patients with lowered WBC (<10,000 per L), 61% (310/503) of patients treated with chemotherapy plus cetuximab responded as compared to 56% (386/695) without cetuximab (OR=1.29, 95% CI=1.02–1.62, P=0.04; $P_{Interaction}$=0.01, Figure 5.4).

5.3.2 Statistical power for the genome wide association study

Based on a sample size of 319 patients (treated with cetuximab and having with *RAS* wild- type tumours) and a MAF≥20%, the univariate analysis had 80% power to detect variants with ORs ≥4.1 and ≥6.2 at suggestive ($1x10^{-5}$) and genome-wide significance ($5x10^{-8}$), respectively. The multivariate GWAS analysis (n=309 patients) had 80% power to detect variants with ORs ≥4.3 at suggestive significance and ≥6.5 at genome-wide significance, (Table 5.3). The reported MAF for these calculations is 20%, which is higher than the usually reported 5% in the rest of this thesis.

5.3.3 Known prognostic factors

Association tests (logistic and Cox regressions) for known prognostic factors were performed on patients with *RAS* mutant tumours treated with cetuximab in order to ensure that multivariate analyses included key prognostic factors. *BRAF* mutation status was significantly associated with response at 12 weeks (OR=0.32, 95% CI=0.15–0.65, P=$1.6x10^{-3}$). Number of metastatic sites (HR=1.63, 95% CI=1.24–2.12, P=$3.7x10^{-4}$), white cell count (HR=1.07, 95% CI=1.04–1.11, P=$5.3x10^{-6}$) and *BRAF* mutation status (HR=3.10, 95% CI=1.50–6.41, P=$2.2x10^{-3}$) are significantly associated with OS (Table 5.2).

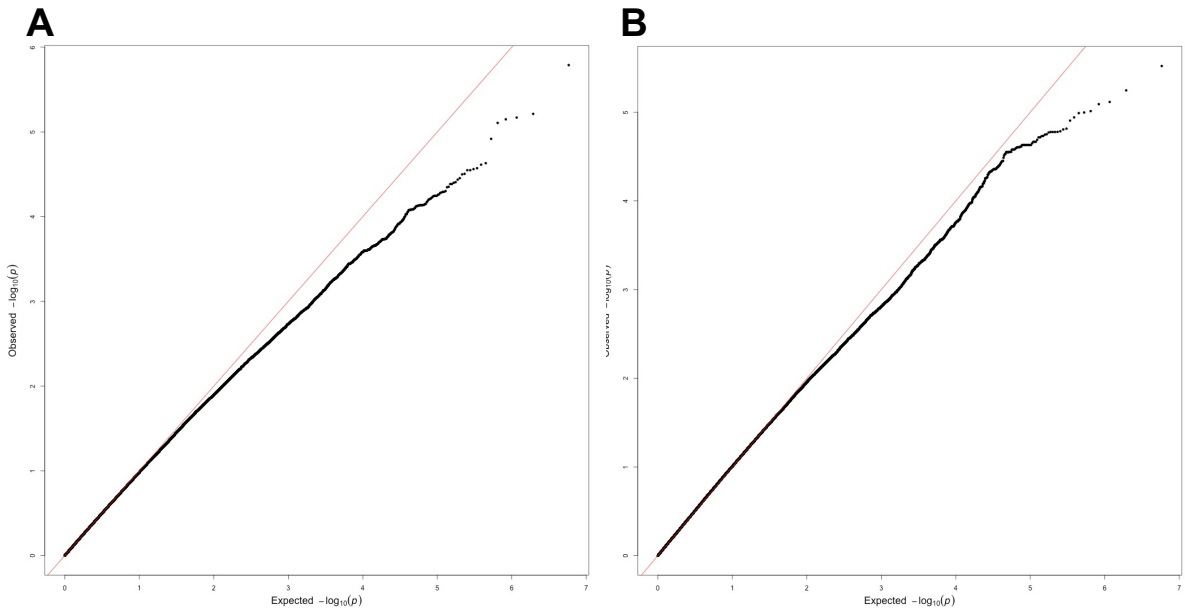

Figure 5.5. Q-Q plot of observed vs expected P values for response to cetuximab in (A) univariate (n=319) and (B) multivariate additive model (n=309)

Genomic inflation factor (λ) univariate=1.01; multivariate=1.02.

5.3.4 Genomic inflation and Population Substructure

Q-Q plots of observed versus expected *P*-values (Figure 5.5) showed no evidence for an inflation of test statistics in either the univariate (λ=1.01) or multivariate (λ=1.02) GWAS for response to cetuximab.

5.3.5 Univariate GWAS for response to cetuximab on *RAS* wild type patients

No SNPs were associated with response to cetuximab at genome-wide significant levels. Two loci, defined by the lead SNPs rs12054810 at 5q23.3 and rs73200904 at 13q21.32, were suggestive of association with response to cetuximab ($P<1x10^{-5}$; Figure 5.6, Table 5.4).

5.3.6 Multivariate GWAS for response to cetuximab on *RAS* wild type patients

No SNPs were associated with response to cetuximab at genome-wide significance levels. The two loci suggestive of association with response to cetuximab in the univariate analysis – rs12054810 at 5q23.3 and rs73200904 at 13q21.32 – were also suggestive of association in the multivariate analysis (rs12054810 multivariate OR=0.37, 95% CI=0.24–0.56, $P=3.0x10^{-6}$; rs73200904 multivariate OR=0.38, 95% CI=0.25–0.58, $P=5.7x10^{-6}$).

Two further loci were suggestive of association with cetuximab in the multivariate analysis that were not in the univariate analyses – rs142144203 at 10q13.31 and rs131850 at 22q23.31 (Figure 5.7, Table 5.6). All four SNPs had a significant interaction with cetuximab and did not influence OS regardless of *RAS* mutational status (Table 5.5, Figure 5.7). Response rates for the four loci found in the univariate and multivariate models are shown in Table 5.7.

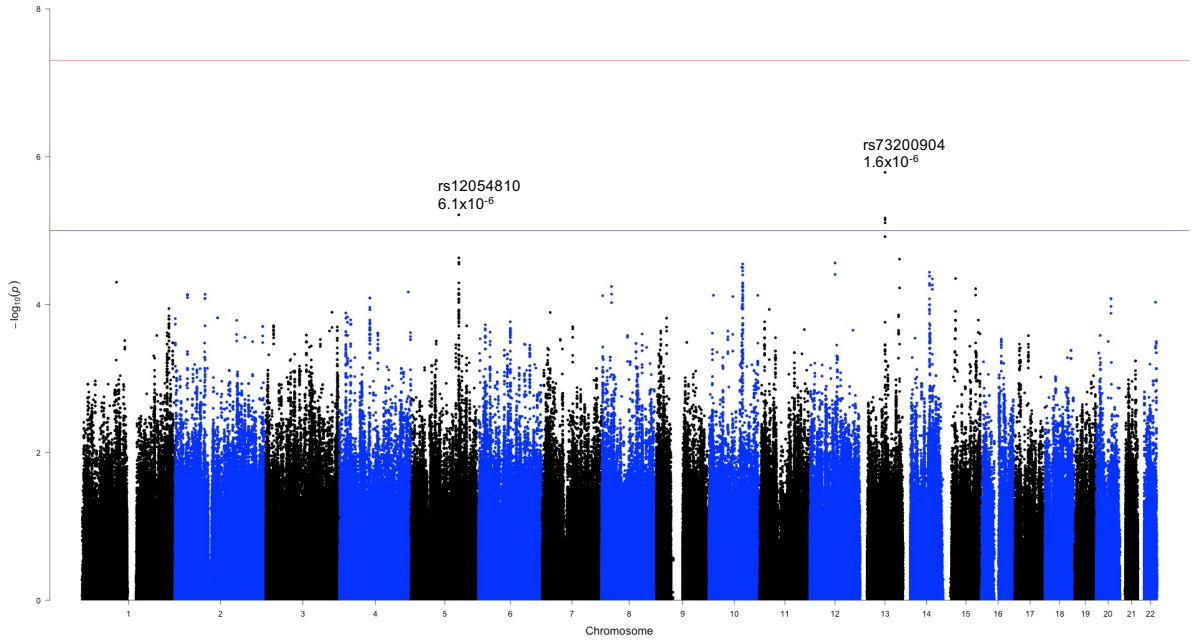

Figure 5.6. Manhattan plot of the relationship between SNP genotype and response to cetuximab in patients with RAS wild-type CRCs (univariate analyses, n=319)

The red line indicates a genome-wide significance threshold of P=5.0x10^{-8} and the blue line indicates suggestive significance threshold of P=1.0x10^{-5}.

Table 5.4. Common variants suggestive of association with response to cetuximab (univariate analyses)

Cytoband	SNP	MAF (%)	Alleles REF/ ALT	OR (95% CI)	*P*	*P* No cet	$P_{Interaction}$	*P* surv
13q21.32	rs73200904	36.5	C/T	0.38 (0.25–0.56)	1.6x10^{-6}	0.67	6.8x10^{-4}	0.54
5q23.3	rs12054810	33.0	A/G	0.41 (0.28–0.60)	6.1x10^{-6}	0.57	2.2x10^{-4}	0.36

Lead SNPs of each independent loci shown. BP=Base Position, MAF=Minor allele frequency, REF=Reference allele, ALT=Alternative allele, OR=Odds ratio, CI=Confidence interval. *P*=*P*-value. *P* no cet= *P*-value for response in patients with *RAS* wild-type CRCs who did not receive cetuximab (n=475), *P* surv=*P*-value for OS.

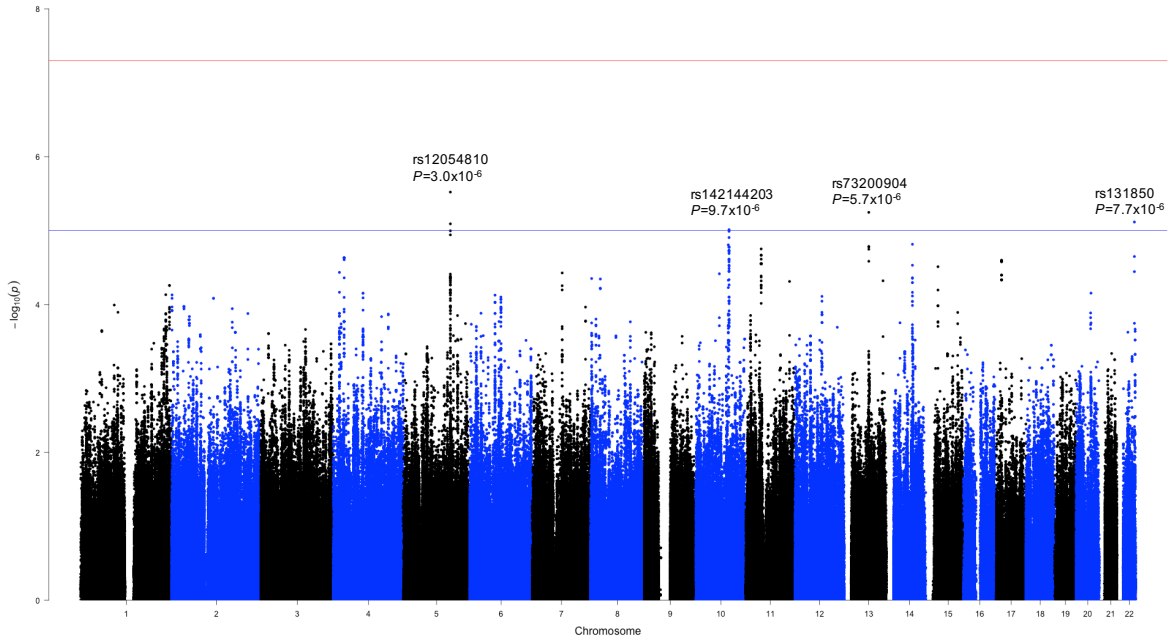

Figure 5.7. Manhattan plot of the relationship between SNP genotype and response to cetuximab in patients with RAS wild-type CRCs (multivariate analyses)

The red line indicates a genome-wide significance threshold of P=5.0x10^{-8} and the blue line indicates suggestive significance threshold of P=1.0x10^{-5}.

Table 5.5. Common variants suggestive of association with response to cetuximab (multivariate analyses)

Cytoband	SNP	MAF (%)	Alleles REF/ ALT	OR (95% CI)	*P*	*P* No cet	$P_{Interaction}$	*P* surv
5q23.3	rs12054810	33.0	A/G	0.37 (0.24–0.56)	3.0x10^{-6}	0.78	2.1x10^{-4}	0.13
13q21.32	rs73200904	36.5	C/T	0.38 (0.25–0.58)	5.7x10^{-6}	0.67	1.5x10^{-3}	0.31
22q23.31	rs131850	39.2	T/C	2.82 (1.79–4.46)	7.7x10^{-6}	0.03	0.02	0.13
10q13.31	rs142144203	27.1	T/-	0.38 (0.24–0.58)	9.7x10^{-6}	0.74	2.8x10^{-4}	0.15

Lead SNPs of independent loci shown. BP=Base Position, MAF=Minor allele frequency, REF=Reference allele, ALT=Alternative allele (-=deletion), OR=Odds ratio, CI=Confidence interval. *P*=*P*-value. *P* no cet=*P*-value for response in patients with *RAS* wild-type CRCs who did not receive cetuximab (n=454), *P* surv=*P*-value for OS. *P*-values were adjusted for age, sex, disease site, number of metastatic sites, World Health Organization performance status, primary tumour resection, *BRAF* mutation status, WBC count and chemotherapy type.

5.3.7 Loci suggestive of significance

rs12054810 is an eQTL for *SLC27A6* and *ISOC1*, and rs73200904 is found in an intron of *PCDH9* (Table 5.6, Figure 5.9). rs142144203 is intronic to *RNLS* and an eQTL for *LIPN* and *ANKRD22* (Table 5.7, Figure 5.8 and 5.9).

5.3.8 Gene-based and gene-set analyses

No genes (Table 5.8) or gene-sets (Table 5.9) were significantly associated with response to cetuximab in either the univariate or multivariate models after correction for multiple testing.

Table 5.6. Genes associated with suggestive loci from the GWAS

Cytoband	SNP	Nearby Genes	Genes in eQTL
Univariate			
13q21.32	rs73200904	*PCDH9, PCDH9-AS2, PCDH9-AS3, PCDH9-AS4*	
5q23.3	rs12054810	*FBN2, SLC27A6, ISCO1*	*SLC27A6, ISCO1*
Multivariate			
5q23.3	rs12054810	*FBN2, SLC27A6, ISCO1*	*SLC27A6, ISCO1*
13q21.32	rs73200904	*PCDH9, PCDH9-AS2, PCDH9-AS3, PCDH9-AS4*	
22q23.31	rs131850		
10q13.31	rs142144203	*RNLS, LIPJ, LIPF, LIPK, LIPN, LIPM, ANKRD22*	*LIPN, ANKRD22*

Table 5.7. Response rates by genotype

	Response rate by genotype (%)		
	AA	**AB**	**BB**
rs12054810	81	70	37
rs142144203	77	71	20
rs73200904	85	68	45
rs131850	60	74	90

For rs12054810 A denotes A allele and B denotes G allele, rs142144203 A denotes GT allele and B denotes G allele, rs73200904 A denotes C allele and B denotes T allele, and rs131850 A denotes T allele and B denotes C allele.

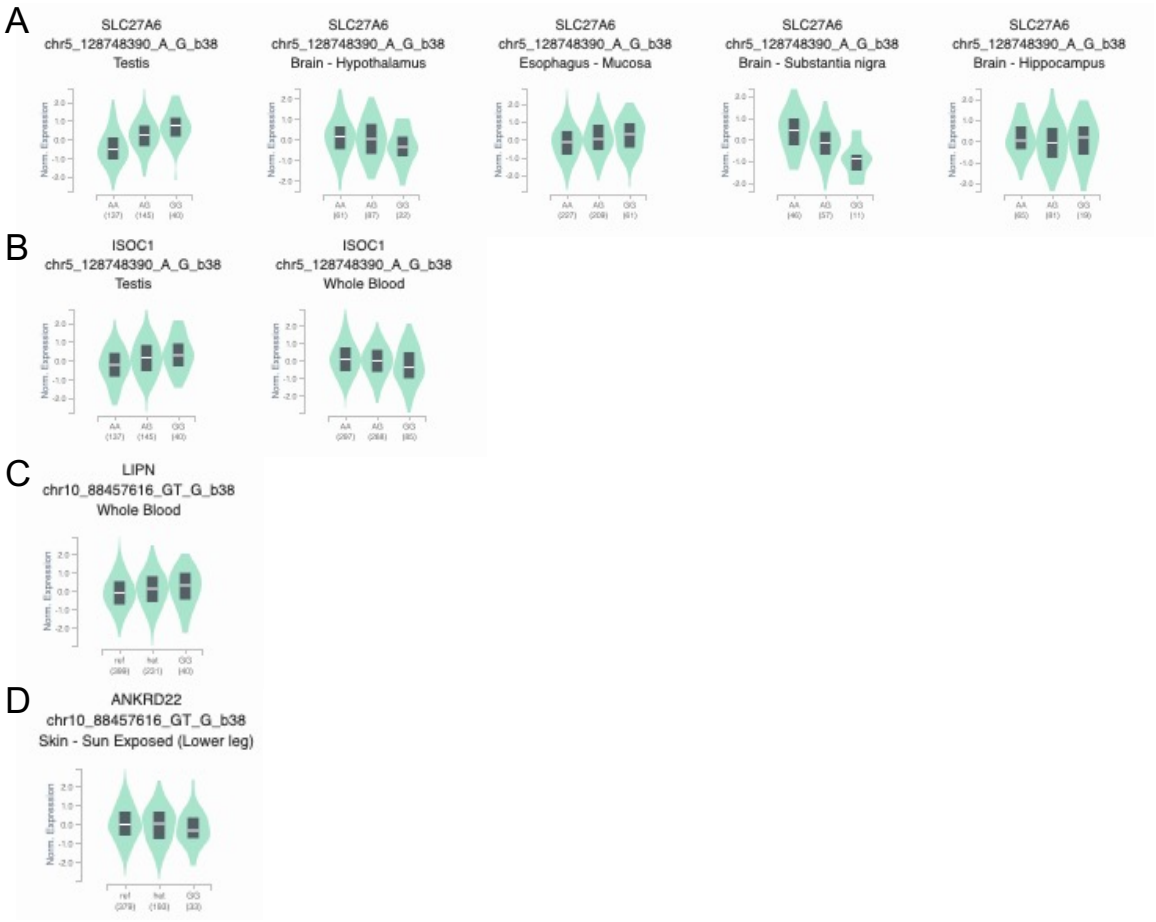

Figure 5.8. Violin plots for eQTL regulation of (A) *SLC27A6* by rs12054810, (B) *ISCO1* by rs12054810 (C) *LIPN* by rs142144203 and (D) *ANKRD22* by rs142144203

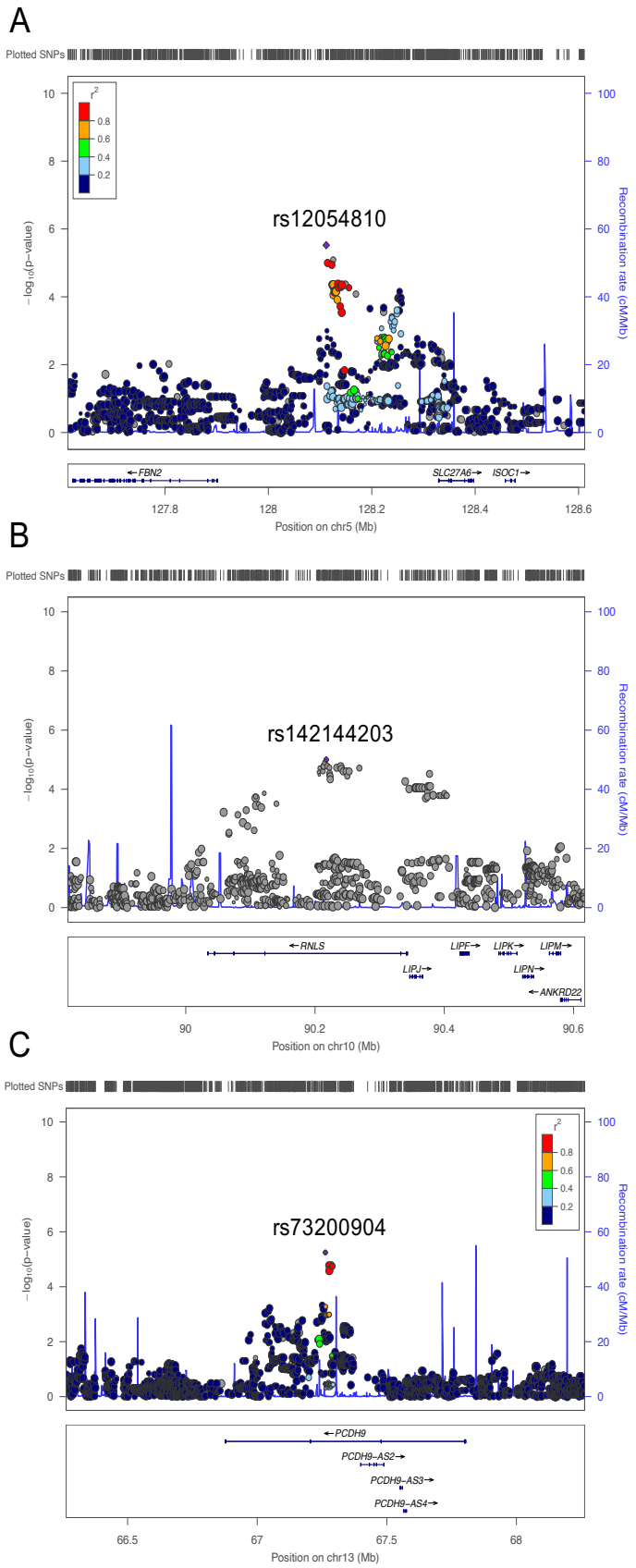

Figure 5.9. Regional plots of (A) 5q23.3, (B) 10q13.31 and (C) 13q21.32 associated with response to cetuximab

Plots show results of the analysis for SNPs and recombination rates. $-\log_{10}(P)$ (y axes) of the SNPs are shown according to their chromosomal positions (x axes). The sentinel SNP (purple) in each analysis is labelled by its rsID. The colour intensity of each symbol reflects the extent of linkage disequilibrium (LD) with the sentinel SNP, deep blue (r^2=0) through to dark red (r^2=1.0) (those in grey lacked LD information). Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a blue line. Physical positions are based on NCBI build 38 of the human genome. Also shown are the relative positions of genes and transcripts mapping to the region of association. Genes have been redrawn to show their relative positions; therefore, maps are not to physical scale.

Table 5.8. Gene-based analysis for response to cetuximab in *RAS* wild-type patients

	Gene ID	Gene Symbol	Gene Name	Chr	SNPs (n)	*P*	*Q*
Univariate							
	5101	*PCDH9*	protocadherin 9	13	1,099	$4.0x10^{-4}$	0.96
	142910	*LIPJ*	lipase family member J	10	39	$7.0x10^{-4}$	0.96
	84324	*SARNP*	SAP domain containing ribonucleoprotein	12	1	$1.0x10^{-3}$	0.96
	29095	*ORMDL2*	ORMDL sphingolipid biosynthesis regulator 2	12	1	$1.0x10^{-3}$	0.96
	23063	*WAPL*	WAPL cohesin release factor	10	161	$1.0x10^{-3}$	0.96
	50852	*TRAT1*	T-cell receptor associated transmembrane adaptor 1	3	125	$1.0x10^{-3}$	0.96
	25960	*ADGRA2*	adhesion G protein-coupled receptor A2	8	3	$1.0x10^{-3}$	0.96
	83752	*LONP2*	lon peptidase 2, peroxisomal	16	261	$2.0x10^{-3}$	0.96
	79090	*TRAPPC6A*	trafficking protein particle complex subunit 6A	19	9	$2.0x10^{-3}$	0.96
	4293	*MAP3K9*	mitogen-activated protein kinase kinase kinase 9	14	298	$2.0x10^{-3}$	0.96
Multivariate							
	142910	*LIPJ*	lipase family member J	10	39	$6.0x10^{-4}$	0.94
	50852	*TRAT1*	T-cell receptor associated transmembrane adaptor 1	3	125	$1.0x10^{-3}$	0.94
	84324	*SARNP*	SAP domain containing ribonucleoprotein	12	1	$1.0x10^{-3}$	0.94
	29095	*ORMDL2*	ORMDL sphingolipid biosynthesis regulator 2	12	1	$1.0x10^{-3}$	0.94
	25960	*ADGRA2*	adhesion G protein-coupled receptor A2	8	3	$1.4x10^{-3}$	0.94

83752	*LONP2*	lon peptidase 2, peroxisomal	16	261	$1.9x10^{-3}$	0.94
79090	*TRAPPC6A*	trafficking protein particle complex subunit 6A	19	9	$1.9x10^{-3}$	0.94
9628	*RGS6*	regulator of G protein signalling 6	14	1,318	$2.2x10^{-3}$	0.94
11212	*PLPBP*	pyridoxal phosphate binding protein	8	6	$2.3x10^{-3}$	0.94
11174	*ADAMTS6*	ADAM metallopeptidase with thrombospondin type 1 motif 6	5	519	$3.9x10^{-3}$	0.94

Univariate GWAS (n=319), Multivariate GWAS (n=309). Statistical significance was corrected for multiple testing using the Bonferroni correction for 20,000 genes to $P=2.5x10^{-6}$. *Q*=*Q*-value to take into account false discovery rate. Ten genes with the smallest *P* values for each GWAS shown.

Table 5.9. Gene-set (pathway) analysis for response to cetuximab in patients with *RAS* wild-type CRCs

	GO ID	GO term	Genes (n)	*Q*
Univariate				
	GO:0006301	Post-replication repair	42	0.55
	GO:0070987	Error-free translesion synthesis	18	0.55
	GO:0033260	Nuclear DNA replication	30	0.55
	GO:0032201	Telomere maintenance via semi-conservative replication	21	0.55
	GO:0071897	DNA biosynthetic process	60	0.55
Multivariate				
	GO:0006595	Polyamine metabolic process	7	0.23
	GO:0009948	Anterior/posterior axis specification	24	0.23
	GO:0008595	Anterior/posterior axis specification, embryo	11	0.23
	GO:0031115	Negative regulation of microtubule polymerization	9	0.23
	GO:0007351	Tripartite regional subdivision	11	0.23

Competitive gene-set analysis performed with all gene results from univariate and multivariate gene analysis. Univariate GWAS (n=319), Multivariate GWAS (n=309). Statistical Significance *Q*=0.05. Q=Q-value to take into account false discovery rate. Five gene-sets with the smallest *Q* values for analysis shown.

5.4 Discussion

5.4.1 Intermittent versus continuous chemotherapy

Regardless of tumoural *RAS* mutation status, the original COIN trial showed non-inferiority of intermittent chemotherapy (when compared to patients given continuous treatment) in patients without cetuximab in randomly allocated treatment groups, or a more stringently selected cohort (Adams et al., 2011). Adams et al. (2011) did however note a modest decrease in survival associated with intermittent treatment (14.4 months) as compared to continuous treatment (15.8 months), although this was not significant (*P* not reported). This is supported by data from the FOCUS4-N trial, which found no significant difference in OS when comparing continuous capecitabine compared to active monitoring (Adams et al., 2021). This investigation explored the relationship between continuous and intermittent oxaliplatin-based chemotherapy by tumoural mutation status and found that patients with *RAS* wild-type CRCs who received continuous chemotherapy alone had better OS (median 20.2 months) as compared to patients who received intermittent chemotherapy alone (median 16.0 months, *P*=0.03). Therefore, this work suggests that patients with *RAS* wild-type CRCs should only receive continuous chemotherapy.

There was no such relationship in patients with *RAS* mutant CRCs (continuous chemotherapy alone, median 16.5 months, intermittent chemotherapy alone, median 16.7 months) possibly due to the poor prognosis associated with *KRAS* mutation status (Yang et al., 2019a). This lack of inferiority of intermittent chemotherapy without cetuximab in patients with *RAS* mutant tumours would allow for treatment breaks and reduce the likelihood of cumulative toxicities which have been shown to reduce quality of life in first-line treatment (Schuurhuizen et al., 2018).

Given that more patients with chemotherapy plus cetuximab had continuous treatment compared to patients without cetuximab, the treatment scheduling benefit

would favour better OS in the group of patients treated with cetuximab. OS could not therefore be used as an accurate marker of cetuximab efficacy in this study.

5.4.2 Tumour biomarkers of response to cetuximab

Response at 12 weeks was used as the primary measure of cetuximab efficacy since at this point patients from the intermittent arms had not undergone any break in treatment. Cetuximab increased response in patients with *RAS* wild-type CRCs from 61% with chemotherapy alone to 72% in those also receiving cetuximab. Despite this improvement in response to cetuximab, the COIN trial showed no improvement in PFS (median 8.6 months in the cetuximab and control groups) or OS (median OS 17.0 versus 17.9 months in the cetuximab and control groups, respectively) (Maughan et al., 2011). The lack of effect on survival was suggested to be caused by a potential interaction with the co-administered chemotherapy, with a benefit associated with fluorouracil-based (FOLFOX) therapy but not with capecitabine-based therapy (XELOX) (Maughan et al., 2011).

This investigation showed a benefit for cetuximab on response in patients treated with XELOX and a trend towards a benefit for those treated with FOLFOX. It, therefore, remains unclear why a difference in response fails to cause a difference in survival outcomes, but this may be due to toxicities associated with the different treatment regimens which can result in dose reduction or treatment delay (Maughan et al., 2011, Wasan et al., 2014).

Despite evidence of association between early response to treatment and OS in this cohort regardless of tumour mutation status (Chapter 2, Section 3.3), there has also shown in some cases to be a ‘disconnection’ between response to therapy and OS, particularly in advanced epithelial cancers (Mittra, 2007, Suzuki et al., 2012). This may be due to, as described by the Gompertz model, the unique growth patterns of tumours (Gompertz, 1825, Winsor, 1932). The Norton-Simon hypothesis, proposed by Simon and Norton (2006) which is based on the Gompertz model, suggests that the rate of death of tumour cells as the result of treatment is proportional to the growth rate of the tumours. Therefore, tumours that respond most quickly to

treatment may also progress more quickly during or after the treatment, resulting in worse prognosis despite good early response rates.

5.4.2.1 Detrimental effect of cetuximab on patients with RAS mutant tumours

In patients with *RAS* mutant CRCs, a detrimental effect of cetuximab on response to therapy, from 54% with chemotherapy alone to 46% in those also receiving cetuximab, was observed in this study. Similar detrimental effects in patients with *RAS* mutant CRCs has been reported in a number of trials including PRIME and CRYSTAL, although the difference is generally not statistically significant (Douillard et al., 2010, Van Cutsem et al., 2011, Zhang et al., 2011). A similar trend was observed in *KRAS* wild-type patients in the original COIN trial, however this was not significant ($P>0.05$) (Maughan et al., 2011). In translational studies, there is also evidence of significantly worse outcome in patients with *KRAS* G12V mutations treated with chemotherapy plus cetuximab compared to treatment with only chemotherapy (Liang et al., 2015). No such detriment has been shown with *KRAS* G13D mutations – where cetuximab has been shown to be more beneficial compared to chemotherapy alone (De Roock et al., 2010b, Mao et al., 2013). This could suggest that the specific mutations in *KRAS* may play a role in resistance to and detriment of cetuximab possibly as a result of different biological responses influenced by the codons mutated within *KRAS*. It has been shown that within CRCs, mutations in *RAS* codon 12 are more associated with mucinous cancers and cancers of advanced stages than tumours with mutations in codon 13 (Bazan et al., 2005, Li et al., 2015). However, this still requires further investigation.

5.4.3 White blood cell count as a predictive biomarker

Raised WBC count predicted worse response to cetuximab at 12 weeks. Interestingly, neutropenia is a potential toxic side effect of cetuximab indicating a possible interaction between changes in white cell count and the treatment (Eli Lilly and Co, 2021).

Some studies have highlighted the influence of specific WBCs on response to cetuximab. Yang et al. (2017a) showed patients with *KRAS* wild-type tumours and high neutrophil count had worse PFS (14.6 vs. 8.12 months, $P<0.01$) and OS (23.0 vs. 13.5 months, $P<0.01$) compared to patients with low neutrophil counts when both groups are treated with cetuximab. The same study showed the opposite effect when investigating lymphocyte counts; with high lymphocyte count resulting in better PFS (14.1 vs 7.0 months, $P<0.01$) and OS (20.0 vs 13.4 months, $P=0.04$). This could show that levels of specific types of WBC or even the ratio of white cell subtypes, as shown in Matić et al. (2017), may be responsible for the significant association with raised WBC and no response to cetuximab in patients with *RAS* wild-type tumours.

5.4.4 Potential germline biomarkers for cetuximab

Despite having excellent tumour-based biomarkers of cetuximab efficacy, there is a lack of germline predictive biomarkers (Chapter 1, Section 3). These GWASs failed to identify any common SNPs associated with response to cetuximab at genome wide significance levels. This may be due to a lack of power to detect modest effect sizes (Table 4.2). SNPs at four loci, two of which were observed in both the univariate and multivariate models, had suggestive associations for response in patients with wild-type tumoural *RAS* and all were predictive for cetuximab; however, none of these influenced OS. The lack of effect on OS may suggest that these biomarkers are false-positives or, as suggested in Section 4.4.2; the tumours responding best to treatment may be more aggressive and therefore not correlated to prognosis.

5.4.4.1 Mechanisms of potential biomarkers

Some genes implicated in this investigation have potential links to the EGFR pathway and therefore are the most biologically significant and more likely to merit further investigation. The two most notable are *PCDH9* – with rs73200904 located in the second intron of this gene – and *ISCO1* which is in eQTL with rs12054810.

Preliminary findings have shown that *PCDH9* may influence the expression of *Matrix metalloproteinase-2* (*MMP2)* and *MMP2* could suppress the EGFR-ERK/AKT signalling pathway (Jiaojiao et al., 2021, Li et al., 2019a). However, the cancer-associated activity of *PCDH9* has mainly been observed in cells from other cancer types including melanoma and glioma (Wang et al., 2014, Jiaojiao et al., 2021). Whilst rs73200904 is found within the intron of *PCDH9*, there is no evidence of this SNP directly influencing expression through an eQTL (Carithers and Moore, 2015). Therefore, further experiments determining the relationship of this SNP and *PCDH9* activity is required to fully understand the mechanism behind this association.

Specifically in CRC cells, knockdown of *ISCO1* has been shown to influence the activity of AKT/GSK-3β pathway found downstream of EGFR and inhibit cellular proliferation and migration (Gao et al., 2019). Therefore upregulation of *ISOC1* associated may upregulate the AKT/GSK-3β pathway resulting in resistance to cetuximab and correlating with the poor response (Carithers and Moore, 2015). However the effect of the minor allele of rs12054810 on *ISCO1* has been shown to differ depending on the tissue (Figure 5.8), with neither tissues significantly associated with differential regulation of this gene found in healthy colon or rectum and therefore, further biological studies, specifically in CRC cells or other resources that provide cancer-specific eQTL data may be required to understand if *ISOC1* is relevant to response to cetuximab in CRC patients.

Based on this evidence, *PCDH9* and *ISOC1* may warrant further investigation as potentially predictive biomarkers for response to cetuximab independent of prognosis.

5.4.5 Gene-based and gene-set analyses

No genes or gene-sets are significantly associated with response to cetuximab in patients with *RAS* wild-type tumours in the univariate or multivariate models either

by using the Bonferroni correction for multiple testing or FDR (*Q*-score) (Bender and Lange, 1999, Bland and Altman, 1995, Wright, 1992).

Interestingly a number of DNA repair processes have been highlighted in gene-sets with the lowest *P*-values, which is supported by DNA repair capacity being linked to response to treatment of CRC (Vodicka et al., 2019). This makes these gene-sets including post-replication repair and DNA biosynthetic process biologically relevant but not statistically significant.

5.4.6 Conclusion and follow up studies

This investigation has confirmed the benefit of cetuximab in patients with *RAS* wild-type CRC. In this group of patients, novel loci suggestive of significant association with response to cetuximab have been identified. Two genes in particular, *PCDH9* and *ISOC1*, have been shown to influence EGFR activity and therefore are important signposts for future investigation.

Despite this, few germline variants were found to be suggestive of association with response to cetuximab in patients with *RAS* wild-type tumours. This may be due to greater influence of tumoural mutations on response to cetuximab, with *BRAF* mutations significantly associated with worse response (P=1.6x10^{-3}). Furthermore, small sample sizes have been shown to limit the power to detect small to medium effect sizes in GWASs (Hong and Park, 2012).

It is important to recognise that the germline variants discovered in this investigation require validation in an independent cohort to be considered genuine biomarkers. Given their odds ratios of 0.37–0.38 and their MAFs of ≥27%, a validation cohort of >100 cases with cetuximab would be required to have sufficient power (80%) to validate these findings (assuming a similar rate of response) at a nominal significance level of P<0.05. This small sample size is due to the large-effect sizes of these variants. Identification of such a validation cohort and meta-analysis with a larger cohort to increase statistical power is a future goal for this work.

6. Chapter 6: General Discussion

This thesis has primarily focused on searching for germline biomarkers for response to therapeutics given in the treatment for aCRC. The main aims were to:

1. Investigate the role of common germline variants (MAF>5%) in response to oxaliplatin-based chemotherapy in the COIN and COIN-B datasets.
2. Investigate the genetic and clinicopathological factors that contribute to treatment outcomes of cetuximab in patients from COIN and COIN-B trials.

6.1 Novel findings

The most notable, novel findings in this thesis are summarised in Table 6.1.

6.1.1 Pattern recognition proteins and oxaliplatin-based chemotherapy

In contrast with previous investigations, this work has been unable to identify associations between OS (or response) and three SNPs in PRPs *FPR1*, *TLR3*, and *TLR4* using multiple statistical models. This negative finding was supported by investigations in a second CRC cohort SCOT, which also found no significant associations with OS and these SNPs (Gray et al., 2019).

6.1.2 Biomarkers for response to oxaliplatin-based chemotherapy

No genome-wide significant associations between germline variants and response to oxaliplatin-based chemotherapy were uncovered as part of this thesis. However, eight loci were found to be suggestive of association. Most notable, SNPs at 10p15.3 were also significantly associated with OS. Interestingly, this locus maps to an area with known links to CRC activity (Zhang et al., 2014b, Jiang et al., 2003, Cornforth et al., 1966, Caruso and Notarnicola, 2005, Virag et al., 2013).

Table 6.1 Notable novel findings in this thesis

Finding	Data	Additional information	Chapter and Section(s)
Germline variants			
Failure to independently validate significant associations with oxaliplatin TOs and SNPs in *FPR1*, *TLR3* and *TLR4*	rs867228, rs3775291, rs3775291 NSD with response or OS	SCOT also failed to validate these findings (Gray et al., 2019)	3.3.5 3.3.6
10p15.3 (rs10903369/ rs2086382) associated with oxaliplatin TOs	↑ Response rs10903369 OR=2.10, P=6.9x10^{-6} ↑ OS rs10903369 HR=0.79, P=3.0x10^{-3}	rs10903369 found within *WDR37* and an eQTL with *IDI1*	4.3.6
Four loci suggestive of association with response to cetuximab	↓ Response rs12054810 OR=0.37, P=3.0x10^{-6}, $P_{Interaction}$=2.1x10^{-4} ↓ Response rs73200904 OR=0.38, P=5.7x10^{-6} $P_{Interaction}$=1.5x10^{-3} ↑ Response rs131850 OR=2.82, P=7.7x10^{-6} $P_{Interaction}$=0.02 ↓ Response rs142144203 OR=0.38, P=7.7x10^{-6} $P_{Interaction}$=2.8x10^{-4}	rs12054810 is an eQTL for *ISOC1* rs73200904 is within the intron of *PCDH9*	5.3.6
Other Factors			
Benefit of continuous chemotherapy for patients with *RAS* wt CRCs	↑ OS HR=0.80, P=0.03	No significant association in *RAS* mut CRCs (P=0.68)	5.2.1.2
RAS mutation status predictive of response to cetuximab	↑ Response RAS wt OR=1.61, P<0.01, $P_{Interaction}$<0.01		5.3.1
WBC associated with OS in patients treated with oxaliplatin	↓ OS high WBC HR=1.64, P=4.0x10^{-13}		4.3.2
WBC associated with response to cetuximab	↓ Response high WBC OR=0.69, P=0.07, $P_{Interaction}$=0.01		5.3.1

Chemotherapy type has no influence on cetuximab response rates	FOLFOX OR=1.53, *P*=0.07 XELOX OR=1.64, *P*=0.02 $P_{Interaction}$=0.82	In patients with *RAS* wt CRCs	5.3.1

↑=increased, ↓=decreased, CRC=colorectal cancer, NSD=no significant difference OS=overall survival, $P_{Interaction}$=*P*-value for interaction with cetuximab, TO=treatment outcome, mut=mutant, wt=wild type. All statistics quoted are for the additive model unless otherwise stated.

6.1.3 Biomarkers for response to cetuximab

In this thesis I have shown that cetuximab improved response in *RAS* wild-type CRCs. This finding is supported by other studies including De Roock et al. (2008). Despite this improved response, investigations of COIN could not find a comparable improvement in PFS or OS and it has been postulated that this lack of survival benefit is due an interaction with the co-administered chemotherapy, XELOX (Maughan et al., 2011). This thesis however, identified a significant benefit for cetuximab on response in patients treated with XELOX and a trend towards a benefit in patients treated with FOLFOX (Chapter 5, Section 3.1). It therefore remains unclear why an increased early response rate does not translate into extended survival (Maughan et al., 2011, Wasan et al., 2014).

Two loci (lead SNPs rs73200904 and rs12054810) were suggestive of association in both models and linked to the EGFR pathway (Gao et al., 2019, Jiaojiao et al., 2021, Li et al., 2019a).

6.1.4 Other reported associations

Raised WBC ($\geq 10^5$/L) was predictive for response to cetuximab in all patients (Chapter 5, Section 3.1). There was also a significant association between WBC and OS in patients with *RAS* wild-type CRCs and treated with cetuximab (Chapter 5, Section 3.3). A similar association between WBC and OS was seen in all patients treated with oxaliplatin (Chapter 4, Section 3.2).

6.2 Implications of the novel findings in this thesis

The detection and validation of biomarkers for treatment outcomes is important for the management of mCRC in the future. New genetic associations for poor treatment outcomes (no response and worse OS) provide the opportunity for better understanding of the molecular pathways for treatment resistance and new targets for personalised therapies (Verdaguer et al., 2017).

New approaches in cancer care towards patient-centered oncology and SDM is increasing. The key objectives of SDM are 1) patients who are fully informed of treatment options including risks and benefits of treatments and 2) patients' preferences and values are incorporated into decision making process (Katz et al., 2014). There is also evidence that patients who are more involved in SDM have better psychological and (in some cases) physical outcomes (Griffin et al., 2004). Therefore, an ability to provide further information on possible outcomes through these newly uncovered predictive biomarkers (once validated) may allow for more informed patients and clinicians and therefore more effective SDM.

6.2.1 Germline biomarkers

This and other investigations have highlighted germline biomarkers for CRC treatment outcomes. It is important that these discovered biomarkers are independently validated in independent cohorts (Oetting et al., 2017). Even when this is the case, no biomarkers for response have made it to the clinic (Morgen et al., 2017). This may be to the modest effect sizes of germline variants in complex diseases (Frazer et al., 2009, Bush and Moore, 2012).

The discordance between the results of Chapter 3 of this thesis and previous studies into *FPR1*, *TLR3* and *TLR4* confirm, as discussed in Battaglin and Lenz (2019), the importance of independent validation of encouraging findings from modestly sized studies, even in those with pre-clinical data to provide a plausible mechanism for an association.

It is unlikely that any independently validated single SNP associated with treatment outcomes such as response and OS will be used a clinical setting in isolation. However, it may be possible for multiple germline SNPs to be combined (perhaps with somatic mutation status) to create a model with effect sizes that are clinically actionable, a process which has had some success in predicting radiotherapy complications (Oh et al., 2017).

Despite the lack of germline variants currently in use routinely in clinical settings, these present a promising area for future research for various CRC phenotypes. Germline variants have been associated with the risk of developing CRC and toxic responses during treatment (Huyghe et al., 2019, Watts et al., 2021). Further epidemiological investigations of CRC may increase our understanding of CRC development and treatment outcomes and uncover variations which lead to more strategies towards personalised medicine approaches for CRC (Zhang et al., 2014a).

6.2.2 *RAS* mutation status

The results of the investigation into associations with response to cetuximab support the current patient inclusion criteria for cetuximab in the UK and USA (National Institute for Health and Care Excellence, 2017, Eli Lilly and Co, 2021). In COIN and COIN-B no benefit for cetuximab was seen in patients with *RAS* mutant CRCs, who are not eligible for treatment with cetuximab according to NICE guidelines (National Institute for Health and Care Excellence, 2017). These are currently overdue for review due to the Covid-19 pandemic and the results of this thesis suggest that no changes to inclusion guidance based on tumoural *RAS* mutation status should be made.

Interestingly, my investigation highlighted a significant benefit of continuous chemotherapy without cetuximab in patients in *RAS* wild-type CRCs and no such benefit in patients with mutant *RAS* CRCs. There is currently no guidance for treatment delivery based tumoral mutation status and therefore with appropriate validation, this could be practice changing (National Institute for Health and Care

Excellence, 2011, National Institute for Health and Care Excellence, 2020a). No significant difference in OS between continuous and intermittent chemotherapy in patients with *RAS* mutant CRCs would indicate that intermittent therapy may an appropriate guideline to reduce cumulative toxicity from oxaliplatin-based chemotherapies which have been shown to impact physical quality of life in patients with mCRC (Schuurhuizen et al., 2018).

6.2.3 Other findings

This investigation has also highlighted a potential predictive effect WBC on response to therapy (oxaliplatin-based chemotherapy with and without cetuximab) and OS. This indicates WBC may be a prognostic factor and with appropriate supportive studies such as Krakowska et al. (2016) or investigation of white cell subtypes, which have also been shown by Watt et al. (2015) to impact CRC outcomes. This shows that WBC should be a focus for further investigation in clinical cohorts and with further supporting results, could be an important tool in personalised CRC treatment.

Current NICE guidelines for cetuximab indicate that cetuximab should be delivered combination with either FOLFOX or FOLFIRI for patients with the appropriate mutational background (National Institute for Health and Care Excellence, 2017). Here, however, I found a benefit for cetuximab on response in patients treated with XELOX and a trend towards a benefit for those treated with FOLFOX which does not support these guidelines.

6.3 Strengths and limitations

The homogeneous nature of response and survival rates in COIN and COIN-B allowed for combination of the two cohorts (Chapter 2, Section 3.3), which becomes a rich, large dataset with detailed genetic and other clinicopathological factors. This cohort and the methodologies applied to the data have strengths and limitations which may influence the findings in this thesis.

6.3.1 Power and statistical significance

The use of the combined COIN and COIN-B cohort for analyses provides sufficient power to detect medium to large effect sizes when examining association for response to oxaliplatin-based chemotherapy and OS (Chapter 3, Section 3.2 and Chapter 4, Section 3.1). Based on power calculations on the combined cohort, only common SNPs (MAF≥5%) were included in GWASs. This is supported by the observation that the majority of associations from published GWASs are in variants with a MAF≥5% and this MAF has appropriate coverage on commercial genotyping arrays (Barrett and Cardon, 2006, Marouli et al., 2017). MAF≥5% has previously been successfully used in GWASs for prognosis and oxaliplatin-based chemotherapy treatment outcomes (including OS) in CRC patients with significant (or suggestive of significance) results (Phipps et al., 2016, Summers, 2019).

Whilst excluding low frequency variants from GWASs reduces the likelihood of type II errors due to insufficient power and less accurate genotyping, the implementation of this threshold will result in low frequency (MAF=1-5%) or rare (MAF<1%) variants with a genuine association with response to treatments being missed.

In contrast, when investigating the potential associations for response to cetuximab in *RAS* wild type patients, a significantly smaller sample (n=319 patients) was used. Therefore the power to detect associations even at MAF>20% is weaker than in the investigation for response to oxaliplatin-based chemotherapy (Chapter 5, Section 3.2). A larger sample size, especially of patients with *RAS* wild-type CRCs treated

with cetuximab, would increase the power to detect smaller effect sizes and reduce the number of false positives (Hirschhorn and Daly, 2005, Biau et al., 2008, Button et al., 2013).

A standard and widely accepted significance threshold of $P<0.05$ was used for the investigation of single variants (Gauvreau and Pagano, 1994). However, in instances when multiple statistical tests are performed, a correction to take this into account was performed. In most instances in this investigation, Bonferroni corrections were employed (Bland and Altman, 1995). It has however been argued that the Bonferroni correction can be overly conservative, so in some instances FDR is also used to try and combat this (Storey et al., 2020, Narum, 2006).

The most notable example of Bonferroni adjusted significance threshold is the genome-wide level of significance ($P<5x10^{-8}$, Bonferroni correction for $P<0.05$ based on one million independent tests) designed to combat the inherent drawback of increased likelihood of false positives in a GWAS (Jannot et al., 2015, Risch and Merikangas, 1996, Phipps et al., 2016). However, due to the inaccurate assumption of independent tests due to SNPs being in LD, this threshold and method is often considered conservative and a possible over-correction (Sham and Purcell, 2014, Fadista et al., 2016, Hirschhorn and Daly, 2005). Despite this, genome-wide significance of $P<5.0x10^{-8}$ is the *de facto* standard significance threshold and has therefore been used in this and many other studies (Ball, 2013, Jannot et al., 2015). In order to overcome this overly conservative significance threshold, the suggestive significance threshold ($P<1x10^{-5}$) was also used in this investigation (Lander and Kruglyak, 1995). Due to the less conservative nature of this threshold, potential biological significance and validation in independent cohorts should be considered when exploring the potential clinical relevance of these variants (Oetting et al., 2017).

6.3.2 Univariate and multivariate investigations

Initially, a univariate analyses of variants was performed which allowed for the investigation of the direct relationship between SNPs and CRC treatment outcomes,

however these analyses do not take into account confounding relationships that may influence the treatment outcomes explored (Ferreira and Purcell, 2009, Kapur, 2017). Using multivariate analyses overcomes this shortcoming, as the method has been shown to yield more informative results in investigations of cancer treatment outcomes such as survival (Bradburn et al., 2003). The addition of biologically and clinically relevant covariates is especially important as tumour mutational status was shown to influence response to both oxaliplatin-based chemotherapy (Chapter 4, Section 3.3) and cetuximab (Chapter 5, Section 3.3).

The strength of the design of COIN and COIN-B is the detailed record of tumour mutational status and other clinical factors which makes multivariate association studies possible. However, this study design also results in some limitations to this multivariate analysis. Validation in independent cohorts may be more difficult if they do not have information on the same or similar clinicopathological factors. Additionally, MSI status, which has shown to be an important biomarker in CRC progression and treatment outcomes, was not included in the multivariate analyses due to limited successful detection of MSI status in the COIN and COIN-B cohorts (Bertagnolli et al., 2009, Lochhead et al., 2013).

Mutations in *KRAS*, *NRAS* and *BRAF* (included in the multivariate analyses) were detected using a combination of pyrosequencing and sequenom of specific known mutational hotspots within these genes (Chapter 2, Section 3.1), but not the entire genes which may result in mutations being missed. Since the sequencing of *C*OIN and COIN-B samples, new methods of sequencing have been adopted including next generation sequencing (NGS), which offers a more comprehensive sequencing approach and can in some cases identify mutations with undefined prognostic and clinical implications (Jones et al., 2017). Whilst there is a high concordance between NGS and standard *KRAS* genotyping, NGS highlighted mutations not tested for on standard *KRAS* assays, which could have potential clinical implications (Kothari et al., 2014). Additionally, Allegra et al. (2016) identified 20% of mCRC patients with tumours originally identified as *KRAS* wild-type that had mutations in exons 3 or 4 of *KRAS* and *NRAS*. Whilst the two genotyping approaches originally employed in COIN and COIN-B were sensitive, the discordance uncovered by Allegra et al. (2016) may have impact on Chapter 5, where a subset of *RAS* wild type patients

were used for the univariate and multivariate GWAS. However, the sequencing methodologies were outside of the scope of this investigation and therefore were not within the candidate's control.

Despite these drawbacks, a combined univariate and multivariate analysis of common SNPs (MAF≥5%) throughout the whole genome resulted in a comprehensive analysis for associations with response to oxaliplatin-based chemotherapy and cetuximab. Whilst there is some consistency between loci suggestive of association with response to oxaliplatin-based chemotherapy or cetuximab examined under univariate and multivariate models in this thesis, they are some loci only found within one model - such as 1q41 which is only found suggestive of response to oxaliplatin-based chemotherapy under a univariate model (Chapter 4, Section 3).

Wang et al. (2017) demonstrated that GWASs which only employ a univariate screening analysis followed by a multivariate analysis on SNPs that are suggestive of significance consistently miss biologically and statistically significant loci which are only highlighted in the multivariate model. This is supported by findings in this thesis, including two loci (lead SNPs rs12054810 and rs131850) suggestive of association with response to cetuximab under a multivariate model that are not seen under a univariate model (Chapter 5, Section 3.6). However, the employment of both methods increases the number of tests performed, which can result in increased false positive results (Ranganathan et al., 2016).

6.3.3 Gene-based and gene-set analyses

The use of gene-based and gene-set analyses using GWAS summary statistics and performed in MAGMA has the ability to extract further biological insights from the in GWASs that are not genome-wide significant and was employed in this thesis (de Leeuw et al., 2015). MAGMA reduces the likelihood of false positives through adjusting the results for confounding variables such as MAF, gene length and SNP density within genes, strengthening the plausibility of the results of these analyses (de Leeuw et al., 2015).

Originally, when exploring genes and gene-sets associated with response to oxaliplatin-based chemotherapy (Chapter 4, Section 3.7), MAGMA version 1.07 was used. However, Yurko et al. (2021) found (in results originally released in 2020) an inflation of gene analysis results, particularly in genes with a large number of SNPs likely due to increased SNP density in modern GWAS summary statistics. This could result in less robust results from investigations using this version, including Chapter 4, Section 3.7. Following the findings from Yurko et al. (2021), a new version of MAGMA (1.08) was released with an adjusted SNP-wise mean model to combat this inflation (de Leeuw et al., 2020). This was used for further analyses including investigations for gene and gene sets associated with response to cetuximab. Whilst this adjustment of methodology is key for statistical accuracy, it is important to note that no genes or gene-sets were found to be significantly associated with response to oxaliplatin-based chemotherapy or cetuximab.

However, despite the implementation of this new version, there are still drawbacks to this methodology (de Leeuw et al., 2020). Some groups of SNPs can still yield inconsistent results when they are in strong LD with each other. Additionally, the role of distal non-coding SNPs on the modulation of expression of a number of genes via the formation of chromatin loops is well established in different cell types (Grundberg et al., 2012, Emmert-Streib et al., 2014, Miele and Dekker, 2009). However, as SNPs are simply assigned to the closest gene by MAGMA's algorithm, this concept is ignored. Whilst there are computational tools to overcome this such as H-MAGMA, they are mainly designed to take into account brain chromatin architecture which would be of limited benefit when investigating the colon or rectum (Sey et al., 2020).

6.4 Future work

6.4.1 Validation of novel biomarkers

The main focus of future work would be an attempt to validate the novel biomarkers uncovered as a result of the work in this thesis. In particular validation of SNPs at 10p15.3 which are suggestive of association with response to oxaliplatin-based chemotherapy and influenced OS in the same cohort (Chapter 4, Section 2) and variants at 5q23.3 and 13q21.32 which are suggestive of association with response to cetuximab in patients with *RAS* wild-type tumours and have possible links to the EGFR pathway (Chapter 5, Section 2). Additionally, data to support the reported benefit of continuous chemotherapy without cetuximab in patients with *RAS* wild type CRCs in COIN and COIN-B would be a future goal as a result of this work.

Where possible, the cohorts used for validation of germline and somatic biomarkers should be stage matched to COIN and COIN-B (stage IV), have similar clinicopathological data and the primary endpoint of radiological response (rather than PFS). This would enable robust validation of suggestive biomarkers for CRC treatment outcomes. The validation cohort may need to be larger than previously discussed (Chapter 4, Section 4.3 and Chapter 5, Section 4.6) due to the winner's curse that is likely to affect the effect sizes reported in this thesis (Oetting et al., 2017, Bigdeli et al., 2016).

6.4.2 Biomarker resequencing

Despite the extensive clinical and molecular data collected as part of the original COIN and COIN-B trials, they employed tumour sequencing methods that are now dated. As discussed in Section 3.3, the pyrosequencing and sequenom used for tumour genotyping have limitations that could be overcome with Next Generation Sequencing (NGS). This largely because NGS has superseded more traditional and conventional molecular biology methodologies (Hert et al., 2008, Behjati and

Tarpey, 2013). It may be interesting and beneficial re-sequence the tumour samples from the patients from COIN and COIN-B.

Re-sequencing would likely result in increased accuracy of genotype calling, ensuring the most accurate *RAS* wild type group for analyses for response to cetuximab. Re-sequencing may also help identify rare(r) mutations and variants in somatic and germline DNA associated with response to aCRC therapeutics. This process which has already been successfully employed in the search for rare variants implicated in schizophrenia (Rhoades et al., 2019). The increased accuracy of genotyping may allow for investigation of effects of MSI as part of the multivariate analyses (which has been shown to influence CRC progression and treatment outcomes) as MSI could not be determined in 37% of genotyped patients (Bertagnolli et al., 2009, Kim et al., 2017, Kawakami et al., 2015). This could be further improved through the use of exome sequencing of regions of interest rather than just hotspots as employed in this investigation (Chapter 2, Section 3.1). This would also allow for further investigation of the role of non-V600E *BRAF* mutant CRCs in resistance to cetuximab which is unclear and comparison of different *KRAS* mutations as there is emerging evidence that some may not be detrimental to cetuximab efficacy (Jones et al., 2017, De Roock et al., 2010b, Mao et al., 2013). Additionally, exome sequencing has been shown to improve the detection of mutations and as a result can enhance access to molecularly targeted therapies in a clinical setting (Réda et al., 2020).

6.4.3 Post-GWAS investigations

6.4.3.1 Investigation of eQTLs

The exploration of eQTLs using the the GTEx database in this thesis has allowed for some investigation of potential gene expression phenotypes associated with SNPs suggestive of association with CRC treatment outcomes, however this in relation to healthy tissue (Nica and Dermitzakis, 2013). Therefore expanding the investigation of potential SNP eQTLS – particularly in cancer cells – is a future goal for this work.

Some resources, such as the human protein atlas can show the difference in gene expression in cancer tissues through staining, this does not demonstrate the relationship between SNPs and expression levels as seen in eQTL analysis (Uhlen et al., 2017). However, the use of publicly available databases with cancer-specific quantitative expression data in relation to SNPs would allow for a more detailed understanding on the effect of SNPs uncovered in this thesis on gene expression.

The use of a resource such a 'PancanQTL' – which uses data from 'The Cancer Genome Atlas' to provide cis-eQTLs, trans-eQTLs, survival-associated eQTLs and GWAS-related eQTLs for over 30 cancer types – would ensure cancer-specific eQTL data (Gong et al., 2018). A main advantage of PancanQTL in particular Importantly for the investigation of COIN and COIN-B patients, both colon and rectal adenocarcinomas are found in this database. Another dataset, the colonomics eQTL browser, enables analyses of eQTL and gene expression in colon tissues. (Moreno et al., 2018). However, samples from 97 colon tumours and 47 healthy colon mucosa were used for the generation of this database so the small sample size must be considered when interpreting results from the colonomics eQTL browser (Moreno et al., 2018). This dataset also does not include any functional information on expression in tissues in rectal cancers, which are studied alongside colon cancers in the COIN and COIN-B clinical trials (Maughan et al., 2011, Wasan et al., 2014).

An alternative method to strengthen this investigation would be to perform mRNA expression analysis of tumour samples from the COIN and COIN-B cohorts. Whilst this is a more expensive approach than the use of publicly available datasets, it provides expression data specific to the patients studies. Combining data from COIN and COIN-B with those found in publicly available databases this would ensure more accurate insights into the eQTLs of SNPs highlighted in this thesis.

6.4.3.2 Other Approaches

Other publicly available databases have been designed to provide further insight into possible functional consequences of SNPs identified during a GWAS (Cano-Gamez and Trynka, 2020). The use of these on the SNPs identified as part of this thesis would allow for a more detailed understanding of the potential biological mechanisms behind the reported associations of CRC treatment outcomes. This is especially applicable for intronic or intergenic SNPs, such as rs131850 (Table 5.6).

For example, SNPs within predicted cis-regulatory elements – regions of non-coding DNA (promoters, enhancers, silencer and operators) that are involved in the regulation of transcription of neighboring genes – influence the behavior of solid tumours such the promotion of transcription of genes involved cell proliferation (Zhou et al., 2020, Tian et al., 2020, MacKenzie et al., 2013). These have been shown to influence gene expression in as response to signal transduction, which is especially relevant when considering the role of EGFR signalling in response to cetuximab (Schoenfelder and Fraser, 2019, MacKenzie et al., 2013). Therefore a future goal is to use of publicly available datasets such as OncoCis – which provides detailed annotation of tissue-specific cis-regulatory genetic variations in cancer – on the results of the GWASs in this thesis (Perera et al., 2014).

Variation in transcription factor bindings sites (TFBS) has also been shown to disrupt gene expression (Maurano et al., 2012, Musunuru et al., 2010). Transcription factors and their binding sites interact with other complexes (including those involved chromatin remodeling) to determine tissue-specific gene expression and the ability of changes in gene expression in response to external signals (Weidemüller et al., 2021, Wiechens et al., 2016, Mullen et al., 2011). Due to the 3D structure of DNA, some gene regulatory sites, including TFBS are rarely near the target gene or within a gene promotor (Mei et al., 2017, Dixon et al., 2012, Schoenfelder and Fraser, 2019). Future exploration of gene regulatory elements may uncover effect of identified SNPs on genes that are not in the same genomic location as the SNPs identified in this thesis. The use of publically available databases such SNP2TFBS would help to achieve this. SNP2TFBS predicts the effects of SNPs on TFBS affinity through position weight matrix model and SEMpl – a SNP effect matrix pipeline

which estimates TFBS affinity through differences in chromatin immunoprecipitation and deep sequencing signal intensity (Kumar et al., 2017, Nishizaki et al., 2020).

The Encyclopedia of DNA Elements (ENCODE) is also a useful tool to provide further insight into results from GWASs through its publicly available functional annotation of the human (and mouse) genome (The ENCODE Project Consortium, 2012). ENCODE is a centralised resource that brings together experimental and computational methods to identify 'biochemically active' regions of the genome and provide a resource for functional characterization and annotation of the genome (Birney et al., 2007, Sloan et al., 2016). Importantly for the investigation of loci identified in this thesis, ENCODE data is also integrated with SNPs, meaning the data can aid in the interpretation of GWAS results (Cano-Gamez and Trynka, 2020, Tulah et al., 2013, Qu and Fang, 2013).

ENCODE provides a large amount of diverse data produced by different of biological and biochemical assays which has been applied to many different cell and tissue types (Moore et al., 2020). These include annotations on DNA methylation found through the reduced representation bisulfite sequencing of human cell lines (Varley et al., 2013). Methylation data from ENCODE information may help with the interpretation of results from this thesis as DNA methylation is a key diagnostic and prognostic marker in cancers such as CRC (Tapial et al., 2019, Hao et al., 2017). Notably, the CpG Islands Methylator Phenotype (CIMP) has also been shown to be important underlying CRC mechanism (Nazemalhosseini Mojarad et al., 2013).

ENCODE can also provide data on the physical interaction of elements of the genome that are not in physical proximity (Grubert et al., 2020). Cohesin-mediated chromatin loops (which facilitates physical interaction of distal elements) are identified via using chromatin interaction analysis by paired-end tag sequencing. A cohesin-mediated chromatin loop has already been shown to play a role in linking a CRC risk-association SNP with upregulation of c-MYC (Wright et al., 2010). Therefore exploring physical interaction of SNPs and other elements in the context of response may provide further insight into the COIN and COIN-B data.

An integrative resource for cancer genomics, EN-CODEC, provides detailed cancer specific annotations from the ENCODE datasets (Zhang et al., 2020). This tool provides detailed and robust annotation (which can be customised) that are experimentally derived from diverse assays. A key aspect of this annotation is comprehensive and experimentally derived networks of both transcription factors and RNA-binding proteins, provides further insights into SNPs and their potential impact on gene expression (Zhang et al., 2020). EN-CODEC can also provide insights on oncogenic transformation and how certain characteristics of cancer genomics change over space and time, vital when considering the role of genetic changes in advances cancers in particular.

Despite the detailed and diverse annotation of the ENCODE resources, there are some drawbacks. For example, it has been argued that the assignment of function to some elements of the genome has been ‘liberal’, with any transcribed element being assigned function (Graur et al., 2013, Kellis et al., 2014). Some of these concerns regarding this assignment of functionality have sincne been addressed (Germain et al., 2014). Despite these concerns, the clear strengths and diversity of data within the resource mean that the use of the ENCODE would be valuable for providing more detailed interpretations of the SNPs highlighted over the course of this thesis (Abascal et al., 2020).

Other databases also provide data on the impact of variants on processes known to be implicated in colorectal cancer including immune infiltration (CancerImmunityQTL), miRNA activity (miRNASNP) and post-transcriptional regulation through alternative polyadenylation (SNP2APA) (Tian et al., 2021, Liu et al., 2020, Yang et al., 2019b, Guo et al., 2020, Gmerek et al., 2019, Mao et al., 2020) The diversity of resources designed for use after GWAS analyses means the opportunity for interpretation is just as diverse. The use of publicly available databases such as these will provide a greater understanding of the functional consequences of loci that have been associated with CRC treatment outcomes. This will ensure the most accurate signposting to appropriate genes, pathways and biological mechanisms related to the SNPs highlighted in this work.

6.4.4 Other approaches

Whilst GWASs have been shown to be successful for identifying new potential germline biomarkers for complex disease phenotypes, such as treatment outcomes (Chapters 4 and 5), they focus on each SNP independently (Szymczak et al., 2009). However, the CD/CV variant suggests that germline variants detected through GWAS will have modest effects and are unlikely to explain more than a small amount of heritability in diseases such as CRC (Frazer et al., 2009). Therefore, measurement of the combined effects of SNPs may lead to more accurate results which can be accountable for more heritability than SNPs alone. An example of this a polygenic risk score (PRS) which combine the weighted effects of many genetic variants into a single score (Lewis and Vassos, 2020). PRSs have been shown to be effective in prediction of risk of developing many multigenic disorders including breast and prostate cancers, diabetes (type I and II) and Alzheimer's disease, with evidence to support their clinical utility within these diseases (Lambert et al., 2019). Based on this evidence, the use of PRS to attempt to predict response to treatments for aCRC may be an appropriate application of the results discussed in this thesis.

The analysis of SNPs simultaneously could be employed using machine learning methodology due to the advances of this field in the context of genetic associations. These methods provide alternatives for GWASs including penalised regression, decision trees and artificial neural network methods (Szymczak et al., 2009). It has been shown that the employment of machine learning models results in improved accuracy of predicting cancer development, recurrence and OS (Cruz and Wishart, 2007). Therefore, an interesting future goal may be to use machine learning methodologies on the COIN and COIN-B genotypes to analyse the combined effects of SNPs on CRC treatment outcomes.

This investigation, through the employment of GWAS, has largely focused on the relationship between single common SNPs (MAF>5%) and CRC treatment outcomes. Most published GWAS results also use a similar threshold, meaning whilst there are some GWASs investigation rare genetic variants, common variants are largely the focus of studies using this methodology (Uffelmann et al., 2021).

However, as stated earlier in this section, the modest effects germline variants detected through GWAS on these phenotypes, more focus on methodologies away from GWAS would allow for further identification of biomarkers for response to treatments for aCRC.

Such methodologies include whole exome sequencing (WES), which sequences all the protein coding exons, which attributes to around 1-3% of the genome (Suwinski et al., 2019, Sakharkar et al., 2004, Guo et al., 2017). This method provides large coverage of genetic variations (including mutations that affect protein coding) with limited cost and storage compared to more comprehensive sequencing methods (Seaby et al., 2016). The use of WES has resulted in the identification of 2 genes (*APCDD1* and *HDAC5*) that may increase the risk of developing familial CRC (Skopelitou et al., 2021). This demonstrates the potential of WES as a tool in the investigation of CRC.

Projects such as the 100,000 genome project, set up by the UK government to apply DNA sequencing to the study of diseases such as a cancer, have demonstrated the power of whole genome sequencing (WGS) in cancer research (Turnbull et al., 2018). Whilst more expensive than WES, it does provide a more even coverage of the genome, rather than just coding sequences (Schwarze et al., 2018, Dunn et al., 2018). Interestingly, it has also been shown that whole genome sequencing has more power to detect exomic sequences than WES (Belkadi et al., 2015).

Pleguezuelos-Manzano et al. (2020), using this rich dataset, have already identified a genotoxic pks+ *E. coli* that causes a ‘distinct mutational signature’ in a subset in colorectal cancers which they believe may serve as a starting point for deeper investigations into the underlying processes of CRC. The use of WGS has also contributed to discovery or rare (and common) genetic risk variants for CRC (Huyghe et al., 2019). The application of WGS to investigations into treatment outcomes in CRC has already identified mutations in *LINC00672* associated with response to treatment (in general) and mutations in *FBXW7* that predict poor response to anti-EGFR treatments (Mendelaar et al., 2021). This study confirms the value of this methodology for investigations into treatment response. Therefore the investigation of response to both cetuximab and oxaliplatin-based chemotherapies

– possibly through re-sequencing of DNA from patients in the COIN and COIN-B trails – via WGS (rather than WES) is a key future aim of this work.

There is also an emerging practice of multi-omic analysis of clinical and population-based datasets. A multi-omic investigation involves the augmentation GWASs with other approaches including proteomics, transcriptomics and epigenomics to provide a more comprehensive and global investigation for associations with a phenotype (Hasin et al., 2017). This has shown to be a promising approach for uncovering as-yet undetected associations in CRC patients, with an unpublished work by Fernandez-Rozadilla et al. (2021) which conducted a transcriptome-wide association study (TWAS) and methylome-wide association study (MWAS) alongside a GWAS and uncovered 103 new risk loci that had not been previously uncovered within a similar cohort. Some of the candidate genes uncovered using this approach had no prior links to CRC tumourigenesis or progression which demonstrates the employment of multi-omics can enhance the possibility of finding novel markers for understanding the development of CRC and signpost for potential therapeutic targets. Based on this success, it may be appropriate to perform similar multi-omics investigations within the COIN and COIN-B (if the expression and methylation data can be extracted from existing samples) to enrich the data discussed in this thesis and ensure a more comprehensive understanding of the molecular mechanisms influencing response in this dataset.

6.5 Outlook

The work in this thesis has confirmed the importance of *RAS* mutations in aCRC. It has also identified novel germline variants suggestive of association with response to oxaliplatin-based chemotherapy (with variations in one locus also significantly associated with overall survival) or cetuximab. Whilst these have yet to be replicated in independent cohorts, they provide a signpost for further investigation which could lead to their use as biomarkers in clinical settings or part of more models, such as polygenic risk scores to provide more detailed predictive models in the future.

7. References

ABASCAL, F., ACOSTA, R., ADDLEMAN, N. J., ADRIAN, J., AFZAL, V., AKEN, B., AKIYAMA, J. A., JAMMAL, O. A., AMRHEIN, H., ANDERSON, S. M., ANDREWS, G. R., ANTOSHECHKIN, I., ARDLIE, K. G., ARMSTRONG, J., ASTLEY, M., BANERJEE, B., BARKAL, A. A., BARNES, I. H. A., BAROZZI, I., BARRELL, D., BARSON, G., BATES, D., BAYMURADOV, U. K., BAZILE, C., BEER, M. A., BEIK, S., BENDER, M. A., BENNETT, R., BOUVRETTE, L. P. B., BERNSTEIN, B. E., BERRY, A., BHASKAR, A., BIGNELL, A., BLUE, S. M., BODINE, D. M., BOIX, C., BOLEY, N., BORRMAN, T., BORSARI, B., BOYLE, A. P., BRANDSMEIER, L. A., BRESCHI, A., BRESNICK, E. H., BROOKS, J. A., BUCKLEY, M., BURGE, C. B., BYRON, R., CAHILL, E., CAI, L., CAO, L., CARTY, M., CASTANON, R. G., CASTILLO, A., CHAIB, H., CHAN, E. T., CHEE, D. R., CHEE, S., CHEN, H., CHEN, H., CHEN, J.-Y., CHEN, S., CHERRY, J. M., CHHETRI, S. B., CHOUDHARY, J. S., CHRAST, J., CHUNG, D., CLARKE, D., CODY, N. A. L., COPPOLA, C. J., COURSEN, J., D'IPPOLITO, A. M., DALTON, S., DANYKO, C., DAVIDSON, C., DAVILA-VELDERRAIN, J., DAVIS, C. A., DEKKER, J., DERAN, A., DESALVO, G., DESPACIO-REYES, G., DEWEY, C. N., DICKEL, D. E., DIEGEL, M., DIEKHANS, M., DILEEP, V., DING, B., DJEBALI, S., DOBIN, A., DOMINGUEZ, D., DONALDSON, S., DRENKOW, J., DRESZER, T. R., DRIER, Y., DUFF, M. O., DUNN, D., EASTMAN, C., ECKER, J. R., EDWARDS, M. D., EL-ALI, N., ELHAJJAJY, S. I., et al. 2020. Perspectives on ENCODE. *Nature,* 583**,** 693-698.

ADAMS, R., BROWN, E., BROWN, L., BUTLER, R., FALK, S., FISHER, D., KAPLAN, R., QUIRKE, P., RICHMAN, S., SAMUEL, L., SELIGMANN, J., SEYMOUR, M., SHIU, K. K., WASAN, H., WILSON, R., MAUGHAN, T. & INVESTIGATORS, F. T. 2018. Inhibition of EGFR, HER2, and HER3 signalling in patients with colorectal cancer wild-type for BRAF, PIK3CA, KRAS, and NRAS (FOCUS4-D): a phase 2-3 randomised trial. *The lancet. Gastroenterology & hepatology,* 3**,** 162-171.

ADAMS, R. A., FISHER, D. J., GRAHAM, J., SELIGMANN, J. F., SEYMOUR, M., KAPLAN, R., YATES, E., PARMAR, M., RICHMAN, S. D., QUIRKE, P., BUTLER, R., BROWN, E., COLLINSON, F., FALK, S., WASAN, H., SHIU, K. K., MIDDLETON, G., SAMUEL, L., WILSON, R. H., BROWN, L. C. & MAUGHAN, T. S. 2021. Capecitabine Versus Active Monitoring in Stable or Responding Metastatic Colorectal Cancer After 16 Weeks of First-Line Therapy: Results of the Randomized FOCUS4-N Trial. *J Clin Oncol***,** Jco2101436.

ADAMS, R. A., MEADE, A. M., MADI, A., FISHER, D., KAY, E., KENNY, S., KAPLAN, R. S., MAUGHAN, T. S., ON BEHALF OF THE, C. T. T. M. G. & INVESTIGATORS 2009. Toxicity associated with combination oxaliplatin plus fluoropyrimidine with or without cetuximab in the MRC COIN trial experience. *British Journal of Cancer,* 100**,** 251-258.

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. & 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**,** 642-53.

ALAZZOUZI, H., ALHOPURO, P., SALOVAARA, R., SAMMALKORPI, H., JÄRVINEN, H., MECKLIN, J. P., HEMMINKI, A., SCHWARTZ, S., JR., AALTONEN, L. A. & ARANGO, D. 2005. SMAD4 as a prognostic marker in colorectal cancer. *Clin Cancer Res,* 11**,** 2606-11.

ALBUQUERQUE, C., BREUKEL, C., VAN DER LUIJT, R., FIDALGO, P., LAGE, P., SLORS, F. J., LEITÃO, C. N., FODDE, R. & SMITS, R. 2002. The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. *Hum Mol Genet,* 11**,** 1549-60.

ALCINDOR, T. & BEAUGER, N. 2011. Oxaliplatin: a review in the era of molecularly targeted therapy. *Curr Oncol,* 18**,** 18-25.

ALLEGRA, C. J., RUMBLE, R. B., HAMILTON, S. R., MANGU, P. B., ROACH, N., HANTEL, A. & 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. *J Clin Oncol,* 34**,** 179-85.

ALTMAN, D. G. & BLAND, J. M. 2005. Standard deviations and standard errors. *BMJ (Clinical research ed.),* 331**,** 903-903.

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. & CHANG, D. D. 2008. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol,* 26**,** 1626-34.

AMARANTE-MENDES, G. P., ADJEMIAN, S., BRANCO, L. M., ZANETTI, L. C., WEINLICH, R. & BORTOLUCI, K. R. 2018. Pattern Recognition Receptors and the Host Cell Death Molecular Machinery. *Front Immunol,* 9**,** 2379.

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. & 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**,** 93-99.

ANAND, P., KUNNUMAKKARA, A. B., SUNDARAM, C., HARIKUMAR, K. B., THARAKAN, S. T., LAI, O. S., SUNG, B. & AGGARWAL, B. B. 2008. Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res,* 25**,** 2097-116.

ANDERSON, C. A., PETTERSSON, F. H., CLARKE, G. M., CARDON, L. R., MORRIS, A. P. & ZONDERVAN, K. T. 2010. Data quality control in genetic case-control association studies. *Nat Protoc,* 5**,** 1564-73.

ANDRÉ, T., BONI, C., MOUNEDJI-BOUDIAF, L., NAVARRO, M., TABERNERO, J., HICKISH, T., TOPHAM, C., ZANINELLI, M., CLINGAN, P., BRIDGEWATER, J., TABAH-FISCH, I. & DE GRAMONT, A. 2004. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med,* 350**,** 2343-51.

ANDRÉ, T., BONI, C., NAVARRO, M., TABERNERO, J., HICKISH, T., TOPHAM, C., BONETTI, A., CLINGAN, P., BRIDGEWATER, J., RIVERA, F. & DE GRAMONT, A. 2009. Improved overall survival with oxaliplatin, fluorouracil,

and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. *J Clin Oncol,* 27**,** 3109-16.

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., LÖVIG, 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. & UROSEVIC, N. 2001. Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study. *Br J Cancer,* 85**,** 692-6.

ANDREYEV, H. J., NORMAN, A. R., CUNNINGHAM, D., OATES, J. R. & CLARKE, P. A. 1998. Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *J Natl Cancer Inst,* 90**,** 675-84.

APETOH, L., GHIRINGHELLI, F., TESNIERE, A., OBEID, M., ORTIZ, C., CRIOLLO, A., MIGNOT, G., MAIURI, M. C., ULLRICH, E., SAULNIER, P., YANG, H., AMIGORENA, S., RYFFEL, B., BARRAT, F. J., SAFTIG, P., LEVI, F., LIDEREAU, R., NOGUES, C., MIRA, J. P., CHOMPRET, A., JOULIN, V., CLAVEL-CHAPELON, F., BOURHIS, J., ANDRE, F., DELALOGE, S., TURSZ, T., KROEMER, G. & ZITVOGEL, L. 2007. Toll-like receptor 4-dependent contribution of the immune system to anticancer chemotherapy and radiotherapy. *Nat Med,* 13**,** 1050-9.

APPS, M. G., CHOI, E. H. & WHEATE, N. J. 2015. The state-of-play and future of platinum drugs. *Endocr Relat Cancer,* 22**,** R219-33.

ARBOUR, N. C., LORENZ, E., SCHUTTE, B. C., ZABNER, J., KLINE, J. N., JONES, M., FREES, K., WATT, J. L. & SCHWARTZ, D. A. 2000. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet,* 25**,** 187-91.

ARGYRIOU, A. A. 2015. Updates on Oxaliplatin-Induced Peripheral Neurotoxicity (OXAIPN). *Toxics,* 3**,** 187-197.

ARMAGHANY, T., WILSON, J. D., CHU, Q. & MILLS, G. 2012. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res,* 5**,** 19-27.

ARNOLD, M., SIERRA, M. S., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2017. Global patterns and trends in colorectal cancer incidence and mortality. *Gut,* 66**,** 683-691.

AUGUSTUS, G. J. & ELLIS, N. A. 2018. Colorectal Cancer Disparity in African Americans: Risk Factors and Carcinogenic Mechanisms. *Am J Pathol,* 188**,** 291-303.

AUSUBEL, F. M. 2005. Are innate immune signaling pathways in plants and animals conserved? *Nat Immunol,* 6**,** 973-9.

AUTON, A., ABECASIS, G. R., ALTSHULER, D. M., DURBIN, R. M., ABECASIS, G. R., BENTLEY, D. R., CHAKRAVARTI, A., CLARK, A. G., DONNELLY, P., EICHLER, E. E., FLICEK, P., GABRIEL, S. B., GIBBS, R. A., GREEN, E. D., HURLES, M. E., KNOPPERS, B. M., KORBEL, J. O., LANDER, E. S., LEE, C., LEHRACH, H., MARDIS, E. R., MARTH, G. T., MCVEAN, G. A., NICKERSON, D. A., SCHMIDT, J. P., SHERRY, S. T., WANG, J., WILSON, R. K., GIBBS, R. A., BOERWINKLE, E., DODDAPANENI, H., HAN, Y., KORCHINA, V., KOVAR, C., LEE, S., MUZNY, D., REID, J. G., ZHU, Y., WANG, J., CHANG, Y., FENG, Q., FANG, X., GUO, X., JIAN, M., JIANG, H., JIN, X., LAN, T., LI, G., LI, J., LI, Y., LIU, S., LIU, X., LU, Y., MA, X., TANG, M., WANG, B., WANG, G., WU, H., WU, R., XU, X., YIN, Y., ZHANG, D., ZHANG, W., ZHAO, J., ZHAO, M., ZHENG, X., LANDER, E. S., ALTSHULER, D. M., GABRIEL, S. B., GUPTA, N., GHARANI, N., TOJI, L. H., GERRY, N. P., RESCH, A. M., FLICEK, P., BARKER, J., CLARKE, L., GIL, L., HUNT, S. E., KELMAN, G., KULESHA, E., LEINONEN, R., MCLAREN, W. M., RADHAKRISHNAN, R., ROA, A., SMIRNOV, D., SMITH, R. E., STREETER, I., THORMANN, A., TONEVA, I., VAUGHAN, B., ZHENG-BRADLEY, X., BENTLEY, D. R., GROCOCK, R., HUMPHRAY, S., JAMES, T., KINGSBURY, Z., LEHRACH, H., SUDBRAK, R., ALBRECHT, M. W., et al. 2015. A global reference for human genetic variation. *Nature,* 526**,** 68-74.

BABA, H., YAMADA, Y., TAKAHARI, D., MATSUMOTO, H., YOSHIDA, K., NAKAMURA, M., YOSHIDA, M., IWAMOTO, S., SHIMADA, K., KOMATSU, Y., SASAKI, Y., SATOH, T., TAKAHASHI, K., MISHIMA, H., MURO, K., WATANABE, M., SAKATA, Y., MORITA, S., SHIMADA, Y. & SUGIHARA, K. 2017. S-1 and oxaliplatin (SOX) plus bevacizumab versus mFOLFOX6 plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer: updated overall survival analyses of the open-label, non-inferiority, randomised phase III: SOFT study. *ESMO open,* 2**,** e000135-e000135.

BAEG, G. H., MATSUMINE, A., KURODA, T., BHATTACHARJEE, R. N., MIYASHIRO, I., TOYOSHIMA, K. & AKIYAMA, T. 1995. The tumour suppressor gene product APC blocks cell cycle progression from G0/G1 to S phase. *Embo j,* 14**,** 5618-25.

BALL, R. D. 2013. Designing a GWAS: power, sample size, and data structure. *Methods Mol Biol,* 1019**,** 37-98.

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. & PIARD, F. 2008. Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res,* 68**,** 8541-6.

BARRETT, J. C. & CARDON, L. R. 2006. Evaluating coverage of genome-wide association studies. *Nat Genet,* 38**,** 659-62.

BASILE, D., GARATTINI, S. K., BONOTTO, M., ONGARO, E., CASAGRANDE, M., CATTANEO, M., FANOTTO, V., DE CARLO, E., LOUPAKIS, F., URBANO, F., NEGRI, F. V., PELLA, N., RUSSANO, M., BRUNETTI, O., SCARTOZZI, M., SANTINI, D., SILVESTRIS, N., CASADEI GARDINI, A., PUZZONI, M., CALVETTI, L., CARDARELLI, N. & APRILE, G. 2017. Immunotherapy for colorectal cancer: where are we heading? *Expert Opin Biol Ther,* 17**,** 709-721.

BASSO, M., STRIPPOLI, A., ORLANDI, A., MARTINI, M., CALEGARI, M. A., SCHINZARI, G., DI SALVATORE, M., CENCI, T., CASSANO, A., LAROCCA, L. M. & BARONE, C. 2013. KRAS mutational status affects oxaliplatin-based chemotherapy independently from basal mRNA ERCC-1 expression in metastatic colorectal cancer patients. *Br J Cancer,* 108**,** 115-20.

BATTAGLIN, F. & LENZ, H. J. 2019. What Should We Do Better? Lessons from Negative Results of a Biomarker Validation Study. *J Natl Cancer Inst,* 111**,** 754-756.

BATTLE, A., BROWN, C. D., ENGELHARDT, B. E. & MONTGOMERY, S. B. 2017. Genetic effects on gene expression across human tissues. *Nature,* 550**,** 204-213.

BAUDINO, T. A. 2015. Targeted Cancer Therapy: The Next Generation of Cancer Treatment. *Curr Drug Discov Technol,* 12**,** 3-20.

BAZAN, V., AGNESE, V., CORSALE, S., CALÒ, V., VALERIO, M. R., LATTERI, M. A., VIENI, S., GRASSI, N., CICERO, G., DARDANONI, G., TOMASINO, R. M., COLUCCI, G., GEBBIA, N. & RUSSO, A. 2005. Specific TP53 and/or Ki-ras mutations as independent predictors of clinical outcome in sporadic colorectal adenocarcinomas: results of a 5-year Gruppo Oncologico dell'Italia Meridionale (GOIM) prospective study. *Ann Oncol,* 16 Suppl 4**,** iv50-55.

BEESLEY, J., SIVAKUMARAN, H., MORADI MARJANEH, M., SHI, W., HILLMAN, K. M., KAUFMANN, S., HUSSEIN, N., KAR, S., LIMA, L. G., HAM, S., MÖLLER, A., CHENEVIX-TRENCH, G., EDWARDS, S. L. & FRENCH, J. D. 2020. eQTL Colocalization Analyses Identify NTN4 as a Candidate Breast Cancer Risk Gene. *The American Journal of Human Genetics,* 107**,** 778-787.

BEHJATI, S. & TARPEY, P. S. 2013. What is next generation sequencing? *Arch Dis Child Educ Pract Ed,* 98**,** 236-8.

BELKADI, A., BOLZE, A., ITAN, Y., COBAT, A., VINCENT, Q. B., ANTIPENKO, A., SHANG, L., BOISSON, B., CASANOVA, J. L. & ABEL, L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A,* 112**,** 5473-8.

BENDER, R. & LANGE, S. 1999. Multiple test procedures other than Bonferroni's deserve wider use. *BMJ (Clinical research ed.),* 318**,** 600-601.

BENSON, A. B., 3RD, VENOOK, A. P., CEDERQUIST, L., CHAN, E., CHEN, Y. J., COOPER, H. S., DEMING, D., ENGSTROM, P. F., ENZINGER, P. C., FICHERA, A., GREM, J. L., GROTHEY, A., HOCHSTER, H. S., HOFFE, S., HUNT, S., KAMEL, A., KIRILCUK, N., KRISHNAMURTHI, S., MESSERSMITH, W. A., MULCAHY, M. F., MURPHY, J. D., NURKIN, S., SALTZ, L., SHARMA, S., SHIBATA, D., SKIBBER, J. M., SOFOCLEOUS, C. T., STOFFEL, E. M., STOTSKY-HIMELFARB, E., WILLETT, C. G., WU, C. S., GREGORY, K. M. & FREEDMAN-CASS, D. 2017. Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw,* 15**,** 370-398.

BENVENUTI, S., SARTORE-BIANCHI, A., DI NICOLANTONIO, F., ZANON, C., MORONI, M., VERONESE, S., SIENA, S. & BARDELLI, A. 2007. Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. *Cancer Res,* 67**,** 2643-8.

BERGEN, S. E. & PETRYSHEN, T. L. 2012. Genome-wide association studies of schizophrenia: does bigger lead to better results? *Curr Opin Psychiatry,* 25**,** 76-82.

BERNTSSON, J., NODIN, B., EBERHARD, J., MICKE, P. & JIRSTRÖM, K. 2016. Prognostic impact of tumour-infiltrating B cells and plasma cells in colorectal cancer. *Int J Cancer,* 139**,** 1129-39.

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. & 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**,** 1814-21.

BIAU, D. J., KERNÉIS, S. & PORCHER, R. 2008. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. *Clin Orthop Relat Res,* 466**,** 2282-8.

BIENZ, M. & HAMADA, F. 2004. Adenomatous polyposis coli proteins and cell adhesion. *Curr Opin Cell Biol,* 16**,** 528-35.

BIGDELI, T. B., LEE, D., WEBB, B. T., RILEY, B. P., VLADIMIROV, V. I., FANOUS, A. H., KENDLER, K. S. & BACANU, S.-A. 2016. A simple yet accurate correction for winner's curse can predict signals discovered in much larger genome scans. *Bioinformatics,* 32**,** 2598-2603.

BIRNEY, E., STAMATOYANNOPOULOS, J. A., DUTTA, A., GUIGÓ, R., GINGERAS, T. R., MARGULIES, E. H., WENG, Z., SNYDER, M., DERMITZAKIS, E. T., STAMATOYANNOPOULOS, J. A., THURMAN, R. E., KUEHN, M. S., TAYLOR, C. M., NEPH, S., KOCH, C. M., ASTHANA, S., MALHOTRA, A., ADZHUBEI, I., GREENBAUM, J. A., ANDREWS, R. M., FLICEK, P., BOYLE, P. J., CAO, H., CARTER, N. P., CLELLAND, G. K., DAVIS, S., DAY, N., DHAMI, P., DILLON, S. C., DORSCHNER, M. O., FIEGLER, H., GIRESI, P. G., GOLDY, J., HAWRYLYCZ, M., HAYDOCK, A., HUMBERT, R., JAMES, K. D., JOHNSON, B. E., JOHNSON, E. M., FRUM, T. T., ROSENZWEIG, E. R., KARNANI, N., LEE, K., LEFEBVRE, G. C., NAVAS, P. A., NERI, F., PARKER, S. C. J., SABO, P. J., SANDSTROM, R., SHAFER, A., VETRIE, D., WEAVER, M., WILCOX, S., YU, M., COLLINS, F. S., DEKKER, J., LIEB, J. D., TULLIUS, T. D., CRAWFORD, G. E., SUNYAEV, S., NOBLE, W. S., DUNHAM, I., DUTTA, A., GUIGÓ, R., DENOEUD, F., REYMOND, A., KAPRANOV, P., ROZOWSKY, J., ZHENG, D., CASTELO, R., FRANKISH, A., HARROW, J., GHOSH, S., SANDELIN, A., HOFACKER, I. L., BAERTSCH, R., KEEFE, D., FLICEK, P., DIKE, S., CHENG, J., HIRSCH, H. A., SEKINGER, E. A., LAGARDE, J., ABRIL, J. F., SHAHAB, A., FLAMM, C., FRIED, C., HACKERMÜLLER, J., HERTEL, J., LINDEMEYER, M., MISSAL, K., TANZER, A., WASHIETL, S., KORBEL, J., EMANUELSSON, O., PEDERSEN, J. S., HOLROYD, N., TAYLOR, R., SWARBRECK, D., MATTHEWS, N., et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447**,** 799-816.

BLAND, J. M. & ALTMAN, D. G. 1995. Multiple significance tests: the Bonferroni method. *Bmj,* 310**,** 170.

BOGAERT, J. & PRENEN, H. 2014. Molecular genetics of colorectal cancer. *Ann Gastroenterol,* 27**,** 9-14.

BOKEMEYER, C., BONDARENKO, I., HARTMANN, J. T., DE BRAUD, F., SCHUCH, G., ZUBEL, A., CELIK, I., SCHLICHTING, M. & 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**,** 1535-1546.

BOKEMEYER, C., BONDARENKO, I., MAKHSON, A., HARTMANN, J. T., APARICIO, J., DE BRAUD, F., DONEA, S., LUDWIG, H., SCHUCH, G., STROH, C., LOOS, A. H., ZUBEL, A. & KORALEWSKI, P. 2009. Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. *J Clin Oncol,* 27**,** 663-71.

BOLAND, P. M. & FAKIH, M. 2014. The emerging role of neoadjuvant chemotherapy for rectal cancer. *J Gastrointest Oncol,* 5**,** 362-73.

BONILLA, F. A. & OETTGEN, H. C. 2010. Adaptive immunity. *Journal of Allergy and Clinical Immunology,* 125**,** S33-S40.

BOSSÉ, Y. & AMOS, C. I. 2018. A Decade of GWAS Results in Lung Cancer. *Cancer Epidemiol Biomarkers Prev,* 27**,** 363-379.

BRADBURN, M. J., CLARK, T. G., LOVE, S. B. & ALTMAN, D. G. 2003. Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit. *British Journal of Cancer,* 89**,** 605-611.

BRANDT, D. S. & CHU, E. 1997. Future challenges in the clinical development of thymidylate synthase inhibitor compounds. *Oncol Res,* 9**,** 403-10.

BRAZIN, K. N., MALLIS, R. J., DAS, D. K., FENG, Y., HWANG, W., WANG, J. H., WAGNER, G., LANG, M. J. & REINHERZ, E. L. 2015. Structural Features of the αβTCR Mechanotransduction Apparatus That Promote pMHC Discrimination. *Front Immunol,* 6**,** 441.

BRENNER, H., KLOOR, M. & POX, C. P. 2014. Colorectal cancer. *Lancet,* 383**,** 1490-1502.

BRIERLEY, J. D., GOSPODAROWICZ, M. K. & WITTEKIND, C. 2017. *TNM classification of malignant tumours*, John Wiley & Sons.

BUSH, W. S. & MOORE, J. H. 2012. Chapter 11: Genome-wide association studies. *PLoS Comput Biol,* 8**,** e1002822.

BUTTON, K. S., IOANNIDIS, J. P. A., MOKRYSZ, C., NOSEK, B. A., FLINT, J., ROBINSON, E. S. J. & MUNAFÒ, M. R. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience,* 14**,** 365-376.

CAMPI, G., CROSTI, M., CONSOGNO, G., FACCHINETTI, V., CONTI-FINE, B. M., LONGHI, R., CASORATI, G., DELLABONA, P. & PROTTI, M. P. 2003. CD4(+) T cells from healthy subjects and colon cancer patients recognize a carcinoembryonic antigen-specific immunodominant epitope. *Cancer Res,* 63**,** 8481-6.

CANCER GENOME ATLAS NETWORK 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature,* 487**,** 330-7.

CANO-GAMEZ, E. & TRYNKA, G. 2020. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics,* 11.

CARIASO, M. & LENNON, G. 2012. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res,* 40**,** D1308-12.

CARITHERS, L. J. & MOORE, H. M. 2015. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank,* 13**,** 307-8.

CARUSO, M. G. & NOTARNICOLA, M. 2005. Biochemical changes of mevalonate pathway in human colorectal cancer. *Anticancer Res,* 25**,** 3393-7.

CASSIDY, J., CLARKE, S., DIAZ-RUBIO, E., SCHEITHAUER, W., FIGER, A., WONG, R., KOSKI, S., RITTWEGER, K., GILBERG, F. & SALTZ, L. 2011. XELOX vs FOLFOX-4 as first-line therapy for metastatic colorectal cancer: NO16966 updated results. *Br J Cancer,* 105**,** 58-64.

CASSIDY, J., TABERNERO, J., TWELVES, C., BRUNET, R., BUTTS, C., CONROY, T., DEBRAUD, F., FIGER, A., GROSSMANN, J., SAWADA, N., SCHÖFFSKI, P., SOBRERO, A., VAN CUTSEM, E. & DÍAZ-RUBIO, E. 2004. XELOX (capecitabine plus oxaliplatin): active first-line therapy for patients with metastatic colorectal cancer. *J Clin Oncol,* 22**,** 2084-91.

CASTRO, F. A., FORSTI, A., BUCH, S., KALTHOFF, H., KRAUSS, C., BAUER, M., EGBERTS, J., SCHNIEWIND, B., BROERING, D. C., SCHREIBER, S., SCHMITT, M., HAMPE, J., HEMMINKI, K. & SCHAFMAYER, C. 2011. TLR-3 polymorphism is an independent prognostic marker for stage II colorectal cancer. *Eur J Cancer,* 47**,** 1203-10.

CATALANO, I., GRASSI, E., BERTOTTI, A. & TRUSOLINO, L. 2019. Immunogenomics of Colorectal Tumors: Facts and Hypotheses on an Evolving Saga. *Trends in Cancer,* 5**,** 779-788.

CERCEK, A., SIEGEL, C. L., CAPANU, M., REIDY-LAGUNES, D. & SALTZ, L. B. 2013. Incidence of chemotherapy-induced amenorrhea in premenopausal women treated with adjuvant FOLFOX for colorectal cancer. *Clin Colorectal Cancer,* 12**,** 163-7.

CERNAT, L., BLAJ, C., JACKSTADT, R., BRANDL, L., ENGEL, J., HERMEKING, H., JUNG, A., KIRCHNER, T. & HORST, D. 2014. Colorectal cancers mimic structural organization of normal colonic crypts. *PLoS One,* 9**,** e104284.

CHANG, C. C., CHOW, C. C., TELLIER, L. C., VATTIKUTI, S., PURCELL, S. M. & LEE, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience,* 4**,** 7.

CHANOCK, S. 2001. Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis Markers,* 17**,** 89-98.

CHEN, D. N., SONG, C. G., YU, K. D., JIANG, Y. Z., YE, F. G. & SHAO, Z. M. 2015. A Prospective Evaluation of the Association between a Single Nucleotide Polymorphism rs3775291 in Toll-Like Receptor 3 and Breast Cancer Relapse. *PLoS One,* 10**,** e0133184.

CHEN, J., XIE, F., CHEN, K., WANG, D., JIANG, H., LI, J., PAN, F., CHEN, S., ZHANG, Y., RUAN, Z., HUANG, H., ZOU, L. & LIANG, H. 2009a. ERCC5 promoter polymorphisms at -763 and +25 predict the response to oxaliplatin-based chemotherapy in patients with advanced colorectal cancer. *Cancer Biol Ther,* 8**,** 1424-30.

CHEN, J.-Q., WU, Y., YANG, H., BERGELSON, J., KREITMAN, M. & TIAN, D. 2009b. Variation in the Ratio of Nucleotide Substitution and Indel Rates across Genomes in Mammals and Bacteria. *Molecular Biology and Evolution,* 26**,** 1523-1531.

CHINTALA, L., VAKA, S., BARANDA, J. & WILLIAMSON, S. K. 2011. Capecitabine versus 5-fluorouracil in colorectal cancer: where are we now? *Oncology Reviews,* 5**,** 129-140.

CHU, E., CALLENDER, M. A., FARRELL, M. P. & SCHMITZ, J. C. 2003. Thymidylate synthase inhibitors as anticancer agents: from bench to bedside. *Cancer Chemother Pharmacol,* 52 Suppl 1**,** S80-9.

CHUNG, D. C. & RUSTGI, A. K. 2003. The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. *Ann Intern Med,* 138**,** 560-70.

CHUNG, K. Y., SHIA, J., KEMENY, N. E., SHAH, M., SCHWARTZ, G. K., TSE, A., HAMILTON, A., PAN, D., SCHRAG, D., SCHWARTZ, L., KLIMSTRA, D. S., FRIDMAN, D., KELSEN, D. P. & SALTZ, L. B. 2005. Cetuximab shows activity in colorectal cancer patients with tumors that do not express the epidermal growth factor receptor by immunohistochemistry. *J Clin Oncol,* 23**,** 1803-10.

CLARKE, C. N. & KOPETZ, E. S. 2015. BRAF mutant colorectal cancer as a distinct subset of colorectal cancer: clinical characteristics, clinical behavior, and response to targeted therapies. *Journal of gastrointestinal oncology,* 6**,** 660-667.

CLEVERS, H. 2006. Wnt/beta-catenin signaling in development and disease. *Cell,* 127**,** 469-80.

CLOSA, A., CORDERO, D., SANZ-PAMPLONA, R., SOLÉ, X., CROUS-BOU, M., PARÉ-BRUNET, L., BERENGUER, A., GUINO, E., LOPEZ-DORIGA, A., GUARDIOLA, J., BIONDO, S., SALAZAR, R. & MORENO, V. 2014. Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis. *Carcinogenesis,* 35**,** 2039-46.

COHEN, S., CARPENTER, G. & KING, L., JR. 1980. Epidermal growth factor-receptor-protein kinase interactions. Co-purification of receptor and epidermal growth factor-enhanced phosphorylation activity. *J Biol Chem,* 255**,** 4834-42.

CONROY, T., HEBBAR, M., BENNOUNA, J., DUCREUX, M., YCHOU, M., LLEDO, G., ADENIS, A., FAROUX, R., REBISCHUNG, C., KOCKLER, L. & DOUILLARD, J. Y. 2010. Quality-of-life findings from a randomised phase-III study of XELOX vs FOLFOX-6 in metastatic colorectal cancer. *Br J Cancer,* 102**,** 59-67.

CORNFORTH, J. W., CORNFORTH, R. H., POPJAK, G. & YENGOYAN, L. 1966. Studies on the biosynthesis of cholesterol. XX. Steric course of decarboxylation of 5-pyrophosphomevalonate and of the carbon to carbon bond formation in the biosynthesis of farnesyl pyrophosphate. *J Biol Chem,* 241**,** 3970-87.

CORVAISIER, M., MOREAU-AUBRY, A., DIEZ, E., BENNOUNA, J., MOSNIER, J. F., SCOTET, E., BONNEVILLE, M. & JOTEREAU, F. 2005. V gamma 9V delta 2 T cell response to colon carcinoma cells. *J Immunol,* 175**,** 5481-8.

COUPS, E. J., MANNE, S. L., MEROPOL, N. J. & WEINBERG, D. S. 2007. Multiple behavioral risk factors for colorectal cancer and colorectal cancer screening status. *Cancer Epidemiol Biomarkers Prev,* 16**,** 510-6.

COUTINHO, A. & GROTHEY, A. 2016. Adjuvant Therapy After Resection of Metastatic Colorectal Cancer: Clear Proof of Principle in This Setting. *Oncology (Williston Park),* 30**,** 50-1, 53.

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. & 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.

CROTTY, S. 2014. T follicular helper cell differentiation, function, and roles in disease. *Immunity,* 41**,** 529-42.

CRUZ, J. A. & WISHART, D. S. 2007. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform,* 2**,** 59-77.

CUNNINGHAM, D., SIROHI, B., PLUZANSKA, A., UTRACKA-HUTKA, B., ZALUSKI, J., GLYNNE-JONES, R., KORALEWSKI, P., BRIDGEWATER, J., MAINWARING, P., WASAN, H., WANG, J. Y., SZCZYLIK, C., CLINGAN, P., CHAN, R. T., TABAH-FISCH, I. & CASSIDY, J. 2009. Two different first-line 5-fluorouracil regimens with or without oxaliplatin in patients with metastatic colorectal cancer. *Ann Oncol,* 20**,** 244-50.

CUSTODIO, A., MORENO-RUBIO, J., APARICIO, J., GALLEGO-PLAZAS, J., YAYA, R., MAUREL, J., RODRÍGUEZ-SALAS, N., BURGOS, E., RAMOS, D., CALATRAVA, A., ANDRADA, E., DÍAZ-LÓPEZ, E., SÁNCHEZ, A., MADERO, R., CEJAS, P. & FELIU, J. 2014. Pharmacogenetic predictors of outcome in patients with stage II and III colon cancer treated with oxaliplatin and fluoropyrimidine-based adjuvant chemotherapy. *Mol Cancer Ther,* 13**,** 2226-37.

CUTSEM, E. V., LANG, I., FOLPRECHT, G., NOWACKI, M., BARONE, C., SHCHEPOTIN, I., MAUREL, J., CUNNINGHAM, D., CELIK, I. & KOHNE, C. 2010. Cetuximab plus FOLFIRI: Final data from the CRYSTAL study on the association of KRAS and BRAF biomarker status with treatment outcome. *Journal of Clinical Oncology,* 28**,** 3570-3570.

DAHLIN, A. M., PALMQVIST, R., HENRIKSSON, M. L., JACOBSSON, M., EKLÖF, V., RUTEGÅRD, J., OBERG, A. & 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**,** 1845-55.

DANIELSEN, S. A., EIDE, P. W., NESBAKKEN, A., GUREN, T., LEITHE, E. & LOTHE, R. A. 2015. Portrait of the PI3K/AKT pathway in colorectal cancer. *Biochim Biophys Acta,* 1855**,** 104-21.

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. & BONETTI, A. 2000. Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J Clin Oncol,* 18**,** 2938-47.

DE GROOT, P. M., WU, C. C., CARTER, B. W. & MUNDEN, R. F. 2018. The epidemiology of lung cancer. *Transl Lung Cancer Res,* 7**,** 220-233.

DE LEEUW, C., SEY, N. Y. A., POSTHUMA, D. & WON, H. 2020. A response to Yurko et al: H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives. *bioRxiv***,** 2020.09.25.310722.

DE LEEUW, C. A., MOOIJ, J. M., HESKES, T. & POSTHUMA, D. 2015. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology,* 11**,** e1004219-e1004219.

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. & 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. *Lancet Oncol,* 11**,** 753-62.

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. & TEJPAR, S. 2010b. Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab. *Jama,* 304**,** 1812-20.

DE ROOCK, W., PIESSEVAUX, H., DE SCHUTTER, J., JANSSENS, M., DE HERTOGH, G., PERSONENI, N., BIESMANS, B., VAN LAETHEM, J. L., PEETERS, M., HUMBLET, Y., VAN CUTSEM, E. & TEJPAR, S. 2008. KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann Oncol,* 19**,** 508-15.

DEKKER, E., TANIS, P. J., VLEUGELS, J. L. A., KASI, P. M. & WALLACE, M. B. 2019. Colorectal cancer. *The Lancet,* 394**,** 1467-1480.

DELANEAU, O., MARCHINI, J. & ZAGURY, J. F. 2011. A linear complexity phasing method for thousands of genomes. *Nat Methods,* 9**,** 179-81.

DELL'AQUILA, E., CREMOLINI, C., ZEPPOLA, T., LONARDI, S., BERGAMO, F., MASI, G., STELLATO, M., MARMORINO, F., SCHIRRIPA, M., URBANO, F., RONZONI, M., TOMASELLO, G., ZANIBONI, A., RACCA, P., BUONADONNA, A., ALLEGRINI, G., FEA, E., DI DONATO, S., CHIARA, S., TONINI, G., TOMCIKOVA, D., BONI, L., FALCONE, A. & SANTINI, D. 2018. Prognostic and predictive role of neutrophil/lymphocytes ratio in metastatic colorectal cancer: a retrospective analysis of the TRIBE study by GONO. *Ann Oncol,* 29**,** 924-930.

DEMB, J., EARLES, A., MARTÍNEZ, M. E., BUSTAMANTE, R., BRYANT, A. K., MURPHY, J. D., LIU, L. & GUPTA, S. 2019. Risk factors for colorectal cancer significantly vary by anatomic site. *BMJ Open Gastroenterol,* 6**,** e000313.

DESAI, M. 1991. Human development: Concepts and measurement. *European Economic Review,* 35**,** 350-357.

DI NICOLANTONIO, F., MARTINI, M., MOLINARI, F., SARTORE-BIANCHI, A., ARENA, S., SALETTI, P., DE DOSSO, S., MAZZUCCHELLI, L., FRATTINI, M., SIENA, S. & BARDELLI, A. 2008. Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J Clin Oncol,* 26**,** 5705-12.

DIAZ-RUBIO, E., SASTRE, J., ZANIBONI, A., LABIANCA, R., CORTES-FUNES, H., DE BRAUD, F., BONI, C., BENAVIDES, M., DALLAVALLE, G. & HOMERIN, M. 1998. Oxaliplatin as single agent in previously untreated colorectal carcinoma patients: a phase II multicentric study. *Ann Oncol,* 9**,** 105-8.

DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature,* 485**,** 376-380.

DOONG, S. L. & DOLNICK, B. J. 1988. 5-Fluorouracil substitution alters pre-mRNA splicing in vitro. *J Biol Chem,* 263**,** 4467-73.

DOUAIHER, J., RAVIPATI, A., GRAMS, B., CHOWDHURY, S., ALATISE, O. & ARE, C. 2017. Colorectal cancer-global burden, trends, and geographical variations. *J Surg Oncol,* 115**,** 619-630.

DOUBENI, C. A., LAIYEMO, A. O., MAJOR, J. M., SCHOOTMAN, M., LIAN, M., PARK, Y., GRAUBARD, B. I., HOLLENBECK, A. R. & SINHA, R. 2012. Socioeconomic status and the risk of colorectal cancer: an analysis of more than a half million adults in the National Institutes of Health-AARP Diet and Health Study. *Cancer,* 118**,** 3636-44.

DOUILLARD, J. Y., OLINER, K. S., SIENA, S., TABERNERO, J., BURKES, R., BARUGEL, M., HUMBLET, Y., BODOKY, G., CUNNINGHAM, D., JASSEM, J., RIVERA, F., KOCÁKOVA, I., RUFF, P., BŁASIŃSKA-MORAWIEC, M., ŠMAKAL, M., CANON, J. L., ROTHER, M., WILLIAMS, R., RONG, A., WIEZOREK, J., SIDHU, R. & PATTERSON, S. D. 2013. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med,* 369**,** 1023-34.

DOUILLARD, J. Y., SIENA, S., CASSIDY, J., TABERNERO, J., BURKES, R., BARUGEL, M., HUMBLET, Y., BODOKY, G., CUNNINGHAM, D., JASSEM, J., RIVERA, F., KOCÁKOVA, I., RUFF, P., BŁASIŃSKA-MORAWIEC, M., ŠMAKAL, M., CANON, J. L., ROTHER, M., OLINER, K. S., TIAN, Y., XU, F. & SIDHU, R. 2014. Final results from PRIME: randomized phase III study of panitumumab with FOLFOX4 for first-line treatment of metastatic colorectal cancer. *Ann Oncol,* 25**,** 1346-1355.

DOUILLARD, J. Y., SIENA, S., CASSIDY, J., TABERNERO, J., BURKES, R., BARUGEL, M., HUMBLET, Y., BODOKY, G., CUNNINGHAM, D., JASSEM, J., RIVERA, F., KOCÁKOVA, I., RUFF, P., BŁASIŃSKA-MORAWIEC, M., ŠMAKAL, M., CANON, J. L., ROTHER, M., OLINER, K. S., WOLF, M. & 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**,** 4697-705.

DUNICAN, D. S., MCWILLIAM, P., TIGHE, O., PARLE-MCDERMOTT, A. & CROKE, D. T. 2002. Gene expression differences between the microsatellite instability (MIN) and chromosomal instability (CIN) phenotypes in colorectal cancer revealed by high-density cDNA array hybridization. *Oncogene,* 21**,** 3253-7.

DUNN, P., ALBURY, C. L., MAKSEMOUS, N., BENTON, M. C., SUTHERLAND, H. G., SMITH, R. A., HAUPT, L. M. & GRIFFITHS, L. R. 2018. Next Generation Sequencing Methods for Diagnosis of Epilepsy Syndromes. *Front Genet,* 9**,** 20.

DURBIN, R. M., ALTSHULER, D., DURBIN, R. M., ABECASIS, G. R., BENTLEY, D. R., CHAKRAVARTI, A., CLARK, A. G., COLLINS, F. S., DE LA VEGA, F. M., DONNELLY, P., EGHOLM, M., FLICEK, P., GABRIEL, S. B., GIBBS, R. A., KNOPPERS, B. M., LANDER, E. S., LEHRACH, H., MARDIS, E. R., MCVEAN, G. A., NICKERSON, D. A., PELTONEN, L., SCHAFER, A. J., SHERRY, S. T., WANG, J., WILSON, R. K., GIBBS, R. A., DEIROS, D., METZKER, M., MUZNY, D., REID, J., WHEELER, D., WANG, J., LI, J., JIAN, M., LI, G., LI, R., LIANG, H., TIAN, G., WANG, B., WANG, J., WANG, W., YANG, H., ZHANG, X., ZHENG, H., LANDER, E. S., ALTSHULER, D.,

AMBROGIO, L., BLOOM, T., CIBULSKIS, K., FENNELL, T. J., GABRIEL, S. B., JAFFE, D. B., SHEFLER, E., SOUGNEZ, C. L., BENTLEY, D. R., GORMLEY, N., HUMPHRAY, S., KINGSBURY, Z., KOKKO-GONZALES, P., STONE, J., MCKERNAN, K. J., COSTA, G. L., ICHIKAWA, J. K., LEE, C. C., SUDBRAK, R., LEHRACH, H., BORODINA, T. A., DAHL, A., DAVYDOV, A. N., MARQUARDT, P., MERTES, F., NIETFELD, W., ROSENSTIEL, P., SCHREIBER, S., SOLDATOV, A. V., TIMMERMANN, B., TOLZMANN, M., EGHOLM, M., AFFOURTIT, J., ASHWORTH, D., ATTIYA, S., BACHORSKI, M., BUGLIONE, E., BURKE, A., CAPRIO, A., CELONE, C., CLARK, S., CONNERS, D., DESANY, B., GU, L., GUCCIONE, L., KAO, K., KEBBEL, A., KNOWLTON, J., LABRECQUE, M., MCDADE, L., MEALMAKER, C., MINDERMAN, M., NAWROCKI, A., NIAZI, F., et al. 2010. A map of human genome variation from population-scale sequencing. *Nature,* 467**,** 1061-1073.

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., SCHÜRMANN, P., DÖRK, 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., et al. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature,* 447**,** 1087-93.

EDIN, S., KAPRIO, T., HAGSTRÖM, J., LARSSON, P., MUSTONEN, H., BÖCKELMAN, C., STRIGÅRD, K., GUNNARSSON, U., HAGLUND, C. & PALMQVIST, R. 2019. The Prognostic Importance of CD20+ B lymphocytes in Colorectal Cancer and the Relation to Other Immune Cell subsets. *Scientific Reports,* 9**,** 19997.

EDIN, S., WIKBERG, M. L., RUTEGARD, J., OLDENBORG, P. A. & PALMQVIST, R. 2013. Phenotypic skewing of macrophages in vitro by secreted factors from colorectal cancer cells. *PLoS One,* 8**,** e74982.

EISENHAUER, E. A., THERASSE, P., BOGAERTS, J., SCHWARTZ, L. H., SARGENT, D., FORD, R., DANCEY, J., ARBUCK, S., GWYTHER, S., MOONEY, M., RUBINSTEIN, L., SHANKAR, L., DODD, L., KAPLAN, R.,

LACOMBE, D. & VERWEIJ, J. 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer,* 45**,** 228-47.

EKLÖF, V., WIKBERG, M. L., EDIN, S., DAHLIN, A. M., JONSSON, B. A., ÖBERG, Å., RUTEGÅRD, J. & PALMQVIST, R. 2013. The prognostic role of KRAS, BRAF, PIK3CA and PTEN in colorectal cancer. *Br J Cancer,* 108**,** 2153-63.

ELI LILLY AND CO. 2021. *ERBITUX® (cetuximab) [package insert]* [Online]. U.S. Food and Drug Administration website. Available: https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/125084s277s280lbl.pdf [Accessed 3rd July 2021].

EMMERT-STREIB, F., DE MATOS SIMOES, R., GLAZKO, G., MCDADE, S., HAIBE-KAINS, B., HOLZINGER, A., DEHMER, M. & CAMPBELL, F. 2014. Functional and genetic analysis of the colon cancer network. *BMC bioinformatics,* 15 Suppl 6**,** S6-S6.

ENGELMANN, B. E., HØGDALL, E., HOLLÄNDER, N. H. & IVESON, T. 2016. TransSCOT – translational research based on the Short Course Oncology Therapy (SCOT) cohort. *Annals of Oncology,* 27.

ERICHSEN, H. C. & CHANOCK, S. J. 2004. SNPs in cancer research and treatment. *Br J Cancer,* 90**,** 747-51.

ESFANDI, F., MOHAMMAD REZAEI, F., TAHERI, M., NABY GOL, M., KHOLGHI OSKOOEI, V., NAMVAR, A. & GHAFOURI-FARD, S. 2019. GAS8 and GAS8-AS1 expression in gastric cancer. *Gastroenterol Hepatol Bed Bench,* 12**,** 322-327.

FADISTA, J., MANNING, A. K., FLOREZ, J. C. & GROOP, L. 2016. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics,* 24**,** 1202-1205.

FAIVRE, S., CHAN, D., SALINAS, R., WOYNAROWSKA, B. & WOYNAROWSKI, J. M. 2003. DNA strand breaks and apoptosis induced by oxaliplatin in cancer cells. *Biochem Pharmacol,* 66**,** 225-37.

FAN, J., ZHU, M., WANG, Y., LI, Z., ZHANG, J., WANG, L., SUN, Q., DAI, J., JIN, G., HU, Z., SHEN, H. & MA, H. 2019. Genome-wide analysis of expression quantitative trait loci identified potential lung cancer susceptibility variants among Asian populations. *Carcinogenesis,* 40**,** 263-268.

FAN, R., HUANG, C. H., LO, S. H., ZHENG, T. & 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.

FANG, J. Y. & RICHARDSON, B. C. 2005. The MAPK signalling pathways and colorectal cancer. *Lancet Oncol,* 6**,** 322-7.

FARON, M., PIGNON, J. P., MALKA, D., BOURREDJEM, A., DOUILLARD, J. Y., ADENIS, A., ELIAS, D., BOUCHÉ, O. & 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**,** 166-76.

FEARNHEAD, N. S., BRITTON, M. P. & BODMER, W. F. 2001. The ABC of APC. *Human Molecular Genetics,* 10**,** 721-733.

FEARON, E. R. 2011. Molecular genetics of colorectal cancer. *Annu Rev Pathol,* 6**,** 479-507.

FEARON, E. R. & VOGELSTEIN, B. 1990. A genetic model for colorectal tumorigenesis. *Cell,* 61**,** 759-67.

FEBBO, P. G., LADANYI, M., ALDAPE, K. D., DE MARZO, A. M., HAMMOND, M. E., HAYES, D. F., IAFRATE, A. J., KELLEY, R. K., MARCUCCI, G., OGINO, S., PAO, W., SGROI, D. C. & BIRKELAND, M. L. 2011. NCCN Task Force report: Evaluating the clinical utility of tumor markers in oncology. *J Natl Compr Canc Netw,* 9 Suppl 5**,** S1-32; quiz S33.

FERNANDEZ-ROZADILLA, C., TIMOFEEVA, M., CHEN, Z., LAW, P., THOMAS, M., SCHMIT, S., DÍEZ-OBRERO, V., HSU, L., FERNANDEZ-TAJES, J., PALLES, C., SHERWOOD, K., BRIGGS, S., SVINTI, V., DONNELLY, K., FARRINGTON, S., BLACKMUR, J., VAUGHAN-SHAW, P., SHU, X.-O., LONG, J., CAI, Q., GUO, X., YINGCHANG, BRODERICK, P., HUYGHE, J., HARRISON, T., CONTI, D., DAMPIER, C., SCHUMACHER, F., MELAS, M., RENNERT, G., OBÓN-SANTACANA, M., MARTÍN-SÁNCHEZ, V., MORATALLA-NAVARRO, F., OH, J. H., KIM, J., JEE, S. H., JUNG, K. J., KWEON, S.-S., SHIN, M.-H., SHIN, A., AHN, Y.-O., KIM, D.-H., OZE, I., WEN, W., MATSUO, K., MATSUDA, K., TANIKAWA, C., REN, Z., GAO, Y.-T., JIA, W.-H., HOPPER, J., JENKINS, M., WIN, A. K., PAI, R., FIGUEIREDO, J., HAILE, R., GALLINGER, S., WOODS, M., NEWCOMB, P., DUGGAN, D., CHEADLE, J. P., KAPLAN, R., MAUGHAN, T. S., KERR, R., KERR, D., KIRAC, I., INVESTIGATORS, C. S., INVESTIGATORS, S. T. T. G. A., BÖHM, J., MECKLIN, J.-P., JOUSILAHTI, P., KNEKT, P., AALTONEN, L., RISSANEN, H., PUKKALA, E., ERIKSSON, J., CAJUSO, T., HÄNNINEN, U., KONDELIN, J., PALIN, K., TANSKANEN, T., RENKONEN-SINISALO, L., ZANKE, B., MÄNNISTÖ, S., ALBANES, D., WEINSTEIN, S., RUIZ-NARVAEZ, E., PALMER, J., BUCHANAN, D., PLATZ, E., VISVANATHAN, K., ULRICH, C., SIEGEL, E., BREZINA, S., GSUR, A., CAMPBELL, P., CHANG-CLAUDE, J., HOFFMEISTER, M., BRENNER, H., SLATTERY, M., et al. 2021. Multi-omic analysis of 100,204 Europeans and Asians identifies 103 new 1 colorectal cancer risk associations and provides insights into disease etiology.

FERREIRA, M. A., GAMAZON, E. R., AL-EJEH, F., AITTOMÄKI, K., ANDRULIS, I. L., ANTON-CULVER, H., ARASON, A., ARNDT, V., ARONSON, K. J., ARUN, B. K., ASSERYANIS, E., AZZOLLINI, J., BALMAÑA, J., BARNES, D. R., BARROWDALE, D., BECKMANN, M. W., BEHRENS, S., BENITEZ, J., BERMISHEVA, M., BIAŁKOWSKA, K., BLOMQVIST, C., BOGDANOVA, N. V., BOJESEN, S. E., BOLLA, M. K., BORG, A., BRAUCH, H., BRENNER, H., BROEKS, A., BURWINKEL, B., CALDÉS, T., CALIGO, M. A., CAMPA, D., CAMPBELL, I., CANZIAN, F., CARTER, J., CARTER, B. D., CASTELAO, J. E., CHANG-CLAUDE, J., CHANOCK, S. J., CHRISTIANSEN, H., CHUNG, W. K., CLAES, K. B. M., CLARKE, C. L., ADLARD, J., AHMED, M., BARWELL, J., BRADY, A., BREWER, C., COOK, J., DAVIDSON, R., DONALDSON, A., EASON, J., EELES, R., EVANS, D. G., GREGORY, H., HANSON, H., HENDERSON, A., HODGSON, S., IZATT, L., KENNEDY, M. J., LALLOO, F., MILLER, C., MORRISON, P. J., ONG, K.-R., PERKINS, J., PORTEOUS, M. E., ROGERS, M. T., SIDE, L. E., SNAPE, K., WALKER, L., HARRINGTON, P. A., ARNOLD, N., AUBER, B., BOGDANOVA-MARKOV, N., BORDE, J., CALIEBE, A., DITSCH, N., DWORNICZAK, B., ENGERT, S., FAUST, U., GEHRIG, A., HAHNEN, E., HAUKE, J., HENTSCHEL, J., HEROLD, N.,

HONISCH, E., JUST, W., KAST, K., LARSEN, M., LEMKE, J., NGUYEN, H. P., NIEDERACHER, D., OTT, C.-E., PLATZER, K., POHL-RESCIGNO, E., RAMSER, J., RHIEM, K., STEINEMANN, D., SUTTER, C., VARON-MATEEVA, R., et al. 2019. Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. *Nature Communications,* 10**,** 1741.

FERREIRA, M. A. & PURCELL, S. M. 2009. A multivariate test of association. *Bioinformatics,* 25**,** 132-3.

FINNER, H. & ROTERS, M. 2002. Multiple hypotheses testing and expected number of type I. 		 errors. *Ann. Statist.,* 30**,** 220-238.

FIOLET, T., SROUR, B., SELLEM, L., KESSE-GUYOT, E., ALLÈS, B., MÉJEAN, C., DESCHASAUX, M., FASSIER, P., LATINO-MARTEL, P., BESLAY, M., HERCBERG, S., LAVALETTE, C., MONTEIRO, C. A., JULIA, C. & TOUVIER, M. 2018. Consumption of ultra-processed foods and cancer risk: results from NutriNet-Santé prospective cohort. *BMJ,* 360**,** k322.

FISCHEL, J. L., FORMENTO, P., CICCOLINI, J., ROSTAGNO, P., ETIENNE, M. C., CATALIN, J. & MILANO, G. 2002. Impact of the oxaliplatin-5 fluorouracil-folinic acid combination on respective intracellular determinants of drug activity. *Br J Cancer,* 86**,** 1162-8.

FISHER, D. 2014. IPDMETAN: Stata module for performing two-stage IPD meta-analysis. S457785\ ed.: Boston College Department of Economics\.

FLANNICK, J., KORN, J. M., FONTANILLAS, P., GRANT, G. B., BANKS, E., DEPRISTO, M. A. & ALTSHULER, D. 2012. Efficiency and power as a function of sequence coverage, SNP array density, and imputation. *PLoS Comput Biol,* 8**,** e1002604.

FLETCHER, R. H. 2009. The diagnosis of colorectal cancer in patients with symptoms: finding a needle in a haystack. *BMC Medicine,* 7**,** 18.

FORSTMEIER, W., WAGENMAKERS, E. J. & PARKER, T. H. 2017. Detecting and avoiding likely false-positive findings - a practical guide. *Biol Rev Camb Philos Soc,* 92**,** 1941-1968.

FOX, J. & SANFORD, W. 2019. An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage.

FOXTROT COLLABORATIVE GROUP 2012. Feasibility of preoperative chemotherapy for locally advanced, operable colon cancer: the pilot phase of a randomised controlled trial. *The Lancet. Oncology,* 13**,** 1152-1160.

FRANK, S. A. 2003. Somatic mutation: early cancer steps depend on tissue architecture. *Curr Biol,* 13**,** R261-3.

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. Y., TSUI, S. K. W., XUE, H., WONG, J. T.-F., 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., SALLÉE, 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., QIANG SONG, Y., TAM, P. K. H., NAKAMURA, Y., KAWAGUCHI, T., KITAMOTO, T., MORIZONO, T., NAGASHIMA, A., OHNISHI, Y., et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature,* 449**,** 851-861.

FRAZER, K. A., MURRAY, S. S., SCHORK, N. J. & TOPOL, E. J. 2009. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics,* 10**,** 241-251.

GALIATSATOS, P. & FOULKES, W. D. 2006. Familial adenomatous polyposis. *Am J Gastroenterol,* 101**,** 385-98.

GALIZIA, G., LIETO, E., DE VITA, F., ORDITURA, M., CASTELLANO, P., TROIANI, T., IMPERATORE, V. & CIARDIELLO, F. 2007. Cetuximab, a chimeric human mouse anti-epidermal growth factor receptor monoclonal antibody, in the treatment of human colorectal cancer. *Oncogene,* 26**,** 3654-60.

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. & PAGES, F. 2006. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science,* 313**,** 1960-4.

GAO, B., ZHAO, L., WANG, F., BAI, H., LI, J., LI, M., HU, X., CAO, J. & WANG, G. 2019. Knockdown of ISOC1 inhibits the proliferation and migration and induces the apoptosis of colon cancer cells through the AKT/GSK-3β pathway. *Carcinogenesis,* 41**,** 1123-1133.

GARCIA-CARBONERO, N., MARTINEZ-USEROS, J., LI, W., ORTA, A., PEREZ, N., CARAMES, C., HERNANDEZ, T., MORENO, I., SERRANO, G. & GARCIA-FONCILLAS, J. 2020. KRAS and BRAF Mutations as Prognostic and Predictive Biomarkers for Standard Chemotherapy Response in Metastatic Colorectal Cancer: A Single Institutional Study. *Cells,* 9.

GAUVREAU, K. & PAGANO, M. 1994. Why 5%? *Nutrition,* 10**,** 93-4.

GENABEL PROJECT DEVELOPERS 2013. GenABEL: genome-wide SNP association analysis. R package version 1.8-0.

GERMAIN, P.-L., RATTI, E. & BOEM, F. 2014. Junk or functional DNA? ENCODE and the function controversy. *Biology & Philosophy,* 29**,** 807-831.

GHIASLOO, M., PAVLENKO, D., VERHAEGHE, M., VAN LANGENHOVE, Z., UYTTEBROEK, O., BERARDI, G., TROISI, R. I. & CEELEN, W. 2020. Surgical treatment of stage IV colorectal cancer with synchronous liver metastases: A systematic review and network meta-analysis. *Eur J Surg Oncol,* 46**,** 1203-1213.

GHIRINGHELLI, F., APETOH, L., TESNIERE, A., AYMERIC, L., MA, Y., ORTIZ, C., VERMAELEN, K., PANARETAKIS, T., MIGNOT, G., ULLRICH, E., PERFETTINI, J. L., SCHLEMMER, F., TASDEMIR, E., UHL, M., GENIN, P., CIVAS, A., RYFFEL, B., KANELLOPOULOS, J., TSCHOPP, J., ANDRE, F., LIDEREAU, R., MCLAUGHLIN, N. M., HAYNES, N. M., SMYTH, M. J., KROEMER, G. & ZITVOGEL, L. 2009. Activation of the

NLRP3 inflammasome in dendritic cells induces IL-1beta-dependent adaptive immunity against tumors. *Nat Med,* 15**,** 1170-8.

GIBBS, R. A., BELMONT, J. W., HARDENBOL, P., WILLIS, T. D., YU, F., YANG, H., CH'ANG, L.-Y., HUANG, W., LIU, B., SHEN, Y., TAM, P. K.-H., TSUI, L.-C., WAYE, M. M. Y., WONG, J. T.-F., ZENG, C., ZHANG, Q., CHEE, M. S., GALVER, L. M., KRUGLYAK, S., MURRAY, S. S., OLIPHANT, A. R., MONTPETIT, A., HUDSON, T. J., CHAGNON, F., FERRETTI, V., LEBOEUF, M., PHILLIPS, M. S., VERNER, A., KWOK, P.-Y., DUAN, S., LIND, D. L., MILLER, R. D., RICE, J. P., SACCONE, N. L., TAILLON-MILLER, P., XIAO, M., NAKAMURA, Y., SEKINE, A., SORIMACHI, K., TANAKA, T., TANAKA, Y., TSUNODA, T., YOSHINO, E., BENTLEY, D. R., DELOUKAS, P., HUNT, S., POWELL, D., ALTSHULER, D., GABRIEL, S. B., ZHANG, H., ZENG, C., MATSUDA, I., FUKUSHIMA, Y., MACER, D. R., SUDA, E., ROTIMI, C. N., ADEBAMOWO, C. A., ANIAGWU, T., MARSHALL, P. A., MATTHEW, O., NKWODIMMAH, C., ROYAL, C. D. M., LEPPERT, M. F., DIXON, M., STEIN, L. D., CUNNINGHAM, F., KANANI, A., THORISSON, G. A., CHAKRAVARTI, A., CHEN, P. E., CUTLER, D. J., KASHUK, C. S., DONNELLY, P., MARCHINI, J., MCVEAN, G. A. T., MYERS, S. R., CARDON, L. R., ABECASIS, G. R., MORRIS, A., WEIR, B. S., MULLIKIN, J. C., SHERRY, S. T., FEOLO, M., ALTSHULER, D., DALY, M. J., SCHAFFNER, S. F., QIU, R., KENT, A., DUNSTON, G. M., KATO, K., NIIKAWA, N., KNOPPERS, B. M., FOSTER, M. W., CLAYTON, E. W., WANG, V. O., WATKIN, J., GIBBS, R. A., BELMONT, J. W., SODERGREN, E., WEINSTOCK, G. M., et al. 2003. The International HapMap Project. *Nature,* 426**,** 789-796.

GMEREK, L., MARTYNIAK, K., HORBACKA, K., KROKOWICZ, P., SCIERSKI, W., GOLUSINSKI, P., GOLUSINSKI, W., SCHNEIDER, A. & MASTERNAK, M. M. 2019. MicroRNA regulation in colorectal cancer tissue and serum. *PLoS One,* 14**,** e0222013.

GOLDSTEIN, N. S. & 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**,** 1331-46.

GOLLINS, S., SUN MYINT, A., HAYLOCK, B., WISE, M., SAUNDERS, M., NEUPANE, R., ESSAPEN, S., SAMUEL, L., DOUGAL, M., LLOYD, A., MORRIS, J., TOPHAM, C. & SUSNERWALA, S. 2011. Preoperative chemoradiotherapy using concurrent capecitabine and irinotecan in magnetic resonance imaging-defined locally advanced rectal cancer: impact on long-term clinical outcomes. *J Clin Oncol,* 29**,** 1042-9.

GOLLINS, S., WEST, N., SEBAG-MONTEFIORE, D., MYINT, A. S., SAUNDERS, M., SUSNERWALA, S., QUIRKE, P., ESSAPEN, S., SAMUEL, L., SIZER, B., WORLDING, J., SOUTHWARD, K., HEMMINGS, G., TINKLER-HUNDAL, E., TAYLOR, M., BOTTOMLEY, D., CHAMBERS, P., LAWRIE, E., LOPES, A. & BEARE, S. 2017. Preoperative chemoradiation with capecitabine, irinotecan and cetuximab in rectal cancer: significance of pre-treatment and post-resection RAS mutations. *Br J Cancer,* 117**,** 1286-1294.

GOMPERTZ, B. 1825. On the nature of the function expressive of the law of human mortality and on the mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society (London)***,** 513-585.

GONG, J., CHO, M. & FAKIH, M. 2016. RAS and BRAF in metastatic colorectal cancer management. *J Gastrointest Oncol,* 7**,** 687-704.

GONG, J., MEI, S., LIU, C., XIANG, Y., YE, Y., ZHANG, Z., FENG, J., LIU, R., DIAO, L., GUO, A. Y., MIAO, X. & HAN, L. 2018. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res,* 46**,** D971-d976.

GRAHAM, J., MUSHIN, M. & KIRKPATRICK, P. 2004. Oxaliplatin. *Nat Rev Drug Discov,* 3**,** 11-2.

GRAHAM, T. A., WEAVER, C., MAO, F., KIMELMAN, D. & XU, W. 2000. Crystal structure of a beta-catenin/Tcf complex. *Cell,* 103**,** 885-96.

GRAUR, D., ZHENG, Y., PRICE, N., AZEVEDO, R. B., ZUFALL, R. A. & ELHAIK, E. 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol,* 5**,** 578-90.

GRAY, R., BARNWELL, J., MCCONKEY, C., HILLS, R. K., WILLIAMS, N. S. & KERR, D. J. 2007. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet,* 370**,** 2020-9.

GRAY, V., BRIGGS, S., PALLES, C., JAEGER, E., IVESON, T., KERR, R., SAUNDERS, M. P., PAUL, J., HARKIN, A., MCQUEEN, J., SUMMERS, M. G., JOHNSTONE, E., WANG, H., GATCOMBE, L., MAUGHAN, T. S., KAPLAN, R., ESCOTT-PRICE, V., AL-TASSAN, N. A., MEYER, B. F., WAKIL, S. M., HOULSTON, R. S., CHEADLE, J. P., TOMLINSON, I. & CHURCH, D. N. 2019. Pattern Recognition Receptor Polymorphisms as Predictors of Oxaliplatin Benefit in Colorectal Cancer. *J Natl Cancer Inst*.

GRIFFIN, S. J., KINMONTH, A. L., VELTMAN, M. W., GILLARD, S., GRANT, J. & STEWART, M. 2004. Effect on health-related outcomes of interventions to alter the interaction between patients and practitioners: a systematic review of trials. *Ann Fam Med,* 2**,** 595-608.

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. & JONES, S. J. 2008. ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res,* 36**,** D107-13.

GRIZZI, F., BASSO, G., BORRONI, E. M., CAVALLERI, T., BIANCHI, P., STIFTER, S., CHIRIVA-INTERNATI, M., MALESCI, A. & LAGHI, L. 2018. Evolving notions on immune response in colorectal cancer and their implications for biomarker development. *Inflamm Res,* 67**,** 375-389.

GRIZZI, F., BIANCHI, P., MALESCI, A. & LAGHI, L. 2013. Prognostic value of innate and adaptive immunity in colorectal cancer. *World J Gastroenterol,* 19**,** 174-84.

GROTHEY, A., VAN CUTSEM, E., SOBRERO, A., SIENA, S., FALCONE, A., YCHOU, M., HUMBLET, Y., BOUCHÉ, O., MINEUR, L., BARONE, C., ADENIS, A., TABERNERO, J., YOSHINO, T., LENZ, H. J., GOLDBERG, R. M., SARGENT, D. J., CIHON, F., CUPIT, L., WAGNER, A. & LAURENT, D. 2013. Regorafenib monotherapy for previously treated metastatic colorectal

cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet,* 381**,** 303-12.
GRUBERT, F., SRIVAS, R., SPACEK, D. V., KASOWSKI, M., RUIZ-VELASCO, M., SINNOTT-ARMSTRONG, N., GREENSIDE, P., NARASIMHA, A., LIU, Q., GELLER, B., SANGHI, A., KULIK, M., SA, S., RABINOVITCH, M., KUNDAJE, A., DALTON, S., ZAUGG, J. B. & SNYDER, M. 2020. Landscape of cohesin-mediated chromatin loops in the human genome. *Nature,* 583**,** 737-743.
GRUNDBERG, E., SMALL, K. S., HEDMAN, Å. K., NICA, A. C., BUIL, A., KEILDSON, S., BELL, J. T., YANG, T.-P., MEDURI, E., BARRETT, A., NISBETT, J., SEKOWSKA, M., WILK, A., SHIN, S.-Y., GLASS, D., TRAVERS, M., MIN, J. L., RING, S., HO, K., THORLEIFSSON, G., KONG, A., THORSTEINDOTTIR, U., AINALI, C., DIMAS, A. S., HASSANALI, N., INGLE, C., KNOWLES, D., KRESTYANINOVA, M., LOWE, C. E., DI MEGLIO, P., MONTGOMERY, S. B., PARTS, L., POTTER, S., SURDULESCU, G., TSAPROUNI, L., TSOKA, S., BATAILLE, V., DURBIN, R., NESTLE, F. O., O'RAHILLY, S., SORANZO, N., LINDGREN, C. M., ZONDERVAN, K. T., AHMADI, K. R., SCHADT, E. E., STEFANSSON, K., SMITH, G. D., MCCARTHY, M. I., DELOUKAS, P., DERMITZAKIS, E. T., SPECTOR, T. D. & MULTIPLE TISSUE HUMAN EXPRESSION RESOURCE, C. 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature genetics,* 44**,** 1084-1089.
GUAN, W. L., QIU, M. Z., HE, C. Y., YANG, L. Q., JIN, Y., WANG, Z. Q., LI, Y. H., XU, R. H. & WANG, F. H. 2020. Clinicopathologic Features and Prognosis of BRAF Mutated Colorectal Cancer Patients. *Front Oncol,* 10**,** 563407.
GULCHER, J. 2012. Microsatellite markers for linkage and association studies. *Cold Spring Harb Protoc,* 2012**,** 425-32.
GUNNLAUGSSON, A., ANDERSON, H., FERNEBRO, E., KJELLÉN, E., BYSTRÖM, P., BERGLUND, K., EKELUND, M., PÅHLMAN, L., HOLM, T., GLIMELIUS, B. & 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**,** 807-13.
GUO, G., GONG, K., WOHLFELD, B., HATANPAA, K. J., ZHAO, D. & HABIB, A. A. 2015. Ligand-Independent EGFR Signaling. *Cancer Res,* 75**,** 3436-41.
GUO, L., WANG, C., QIU, X., PU, X. & CHANG, P. 2020. Colorectal Cancer Immune Infiltrates: Significance in Patient Prognosis and Immunotherapeutic Efficacy. *Frontiers in immunology,* 11**,** 1052-1052.
GUO, Y., DAI, Y., YU, H., ZHAO, S., SAMUELS, D. C. & SHYR, Y. 2017. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics,* 109**,** 83-90.
GUO, Y., XIONG, B. H., ZHANG, T., CHENG, Y. & MA, L. 2016. XELOX vs. FOLFOX in metastatic colorectal cancer: An updated meta-analysis. *Cancer Invest,* 34**,** 94-104.
GUO, Y. A., CHANG, M. M., HUANG, W., OOI, W. F., XING, M., TAN, P. & SKANDERUP, A. J. 2018. Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. *Nature Communications,* 9**,** 1520.
GUREN, T. K., THOMSEN, M., KURE, E. H., SORBYE, H., GLIMELIUS, B., PFEIFFER, P., ÖSTERLUND, P., SIGURDSSON, F., LOTHE, I. M. B.,

DALSGAARD, A. M., SKOVLUND, E., CHRISTOFFERSEN, T. & TVEIT, K. M. 2017. Cetuximab in treatment of metastatic colorectal cancer: final survival analyses and extended RAS data from the NORDIC-VII study. *Br J Cancer,* 116**,** 1271-1278.

HAGGAR, F. A. & BOUSHEY, R. P. 2009. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg,* 22**,** 191-7.

HALAZUN, K. J., ALDOORI, A., MALIK, H. Z., AL-MUKHTAR, A., PRASAD, K. R., TOOGOOD, G. J. & LODGE, J. P. 2008. Elevated preoperative neutrophil to lymphocyte ratio predicts survival following hepatic resection for colorectal liver metastases. *Eur J Surg Oncol,* 34**,** 55-60.

HAMMER, G. E. & MA, A. 2013. Molecular control of steady-state dendritic cell maturation and immune homeostasis. *Annu Rev Immunol,* 31**,** 743-91.

HAMMOND, S. M. 2007. Soaking up small RNAs. *Nature Methods,* 4**,** 694-695.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HANKEY, W., FRANKEL, W. L. & GRODEN, J. 2018. Functions of the APC tumor suppressor protein dependent and independent of canonical WNT signaling: implications for therapeutic targeting. *Cancer Metastasis Rev,* 37**,** 159-172.

HAO, X., LUO, H., KRAWCZYK, M., WEI, W., WANG, W., WANG, J., FLAGG, K., HOU, J., ZHANG, H., YI, S., JAFARI, M., LIN, D., CHUNG, C., CAUGHEY, B. A., LI, G., DHAR, D., SHI, W., ZHENG, L., HOU, R., ZHU, J., ZHAO, L., FU, X., ZHANG, E., ZHANG, C., ZHU, J. K., KARIN, M., XU, R. H. & ZHANG, K. 2017. DNA methylation markers for diagnosis and prognosis of common cancers. *Proc Natl Acad Sci U S A,* 114**,** 7414-7419.

HARBAUM, L., POLLHEIMER, M. J., KORNPRAT, P., LINDTNER, R. A., BOKEMEYER, C. & LANGNER, C. 2015. Peritumoral eosinophils predict recurrence in colorectal cancer. *Modern Pathology,* 28**,** 403-413.

HARRIS, R., BRADBURN, M., DEEKS, J., HARBORD, R., ALTMAN, D. & STERNE, J. 2008. metan: fixed- and random-effects meta-analysis. *Stata Journal,* 8**,** 3-28.

HASIN, Y., SELDIN, M. & LUSIS, A. 2017. Multi-omics approaches to disease. *Genome Biology,* 18**,** 83.

HEINEN, C. D., GOSS, K. H., CORNELIUS, J. R., BABCOCK, G. F., KNUDSEN, E. S., KOWALIK, T. & GRODEN, J. 2002. The APC tumor suppressor controls entry into S-phase through its ability to regulate the cyclin D/RB pathway. *Gastroenterology,* 123**,** 751-63.

HENCKAERTS, L., PIERIK, M., JOOSSENS, M., FERRANTE, M., RUTGEERTS, P. & VERMEIRE, S. 2007. Mutations in pattern recognition receptor genes modulate seroreactivity to microbial antigens in patients with inflammatory bowel disease. *Gut,* 56**,** 1536-42.

HENRIKSEN, L., GRANDAL, M. V., KNUDSEN, S. L., VAN DEURS, B. & GRØVDAL, L. M. 2013. Internalization mechanisms of the epidermal growth factor receptor after activation with different ligands. *PLoS One,* 8**,** e58148.

HENRY, L. & WICKHAM, H. 2020. purrr: Functional Programming Tools. R package version 0.3.4.

HERT, D. G., FREDLAKE, C. P. & BARRON, A. E. 2008. Advantages and limitations of next-generation sequencing technologies: a comparison of

electrophoresis and non-electrophoresis methods. *Electrophoresis,* 29**,** 4618-26.

HIGGINS, G. A., JR., CONN, J. H., JORDAN, P. H., JR., HUMPHREY, E. W., ROSWIT, B. & KEEHN, R. J. 1975. Preoperative radiotherapy for colorectal cancer. *Ann Surg,* 181**,** 624-31.

HILL, W. & ROBERTSON, A. 1968. Linkage disequilibrium in finite populations. *Theoretical and applied genetics,* 38**,** 226-231.

HIRSCHHORN, J. N. & DALY, M. J. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics,* 6**,** 95-108.

HOEVENAAR, W. H. M., JANSSEN, A., QUIRINDONGO, A. I., MA, H., KLAASEN, S. J., TEIXEIRA, A., VAN GERWEN, B., LANSU, N., MORSINK, F. H. M., OFFERHAUS, G. J. A., MEDEMA, R. H., KOPS, G. J. P. L. & JELLUMA, N. 2020. Degree and site of chromosomal instability define its oncogenic potential. *Nature Communications,* 11**,** 1501.

HOFF, P. M., SAAD, E. D., COSTA, F., COUTINHO, A. K., CAPONERO, R., PROLLA, G. & GANSL, R. C. 2012. Literature review and practical aspects on the management of oxaliplatin-associated toxicity. *Clin Colorectal Cancer,* 11**,** 93-100.

HONG, E. P. & PARK, J. W. 2012. Sample size and statistical power calculation in genetic association studies. *Genomics Inform,* 10**,** 117-22.

HOUGHTON, J. A., TILLMAN, D. M. & HARWOOD, F. G. 1995. Ratio of 2'-deoxyadenosine-5'-triphosphate/thymidine-5'-triphosphate influences the commitment of human colon carcinoma cells to thymineless death. *Clin Cancer Res,* 1**,** 723-30.

HOWIE, B. N., DONNELLY, P. & MARCHINI, J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet,* 5**,** e1000529.

HRUSHESKY, W. J., VYZULA, R. & WOOD, P. A. 1999. Fertility maintenance and 5-fluorouracil timing within the mammalian fertility cycle. *Reprod Toxicol,* 13**,** 413-20.

HU, X., QIN, W., LI, S., HE, M., WANG, Y., GUAN, S., ZHAO, H., YAO, W., WEI, M., LIU, M. & WU, H. 2019. Polymorphisms in DNA repair pathway genes and ABCG2 gene in advanced colorectal cancer: correlation with tumor characteristics and clinical outcome in oxaliplatin-based chemotherapy. *Cancer Manag Res,* 11**,** 285-297.

HU, Y., TAO, S. Y., DENG, J. M., HOU, Z. K., LIANG, J. Q., HUANG, Q. G., LI, L. H., LI, H. B., CHEN, Y. M., YI, H., CHEN, X. L. & LIU, H. 2018. Prognostic Value of NRAS Gene for Survival of Colorectal Cancer Patients: A Systematic Review and Meta-Analysis. *Asian Pac J Cancer Prev,* 19**,** 3001-3008.

HUANG, D., SUN, W., ZHOU, Y., LI, P., CHEN, F., CHEN, H., XIA, D., XU, E., LAI, M., WU, Y. & ZHANG, H. 2018. Mutations of key driver genes in colorectal cancer progression and metastasis. *Cancer Metastasis Rev,* 37**,** 173-187.

HUANG, J., HOWIE, B., MCCARTHY, S., MEMARI, Y., WALTER, K., MIN, J. L., DANECEK, P., MALERBA, G., TRABETTI, E., ZHENG, H.-F., AL TURKI, S., AMUZU, A., ANDERSON, C. A., ANNEY, R., ANTONY, D., ARTIGAS, M. S., AYUB, M., BALA, S., BARRETT, J. C., BARROSO, I., BEALES, P., BENN, M., BENTHAM, J., BHATTACHARYA, S., BIRNEY, E., BLACKWOOD, D., BOBROW, M., BOCHUKOVA, E., BOLTON, P. F.,

BOUNDS, R., BOUSTRED, C., BREEN, G., CALISSANO, M., CARSS, K., PABLO CASAS, J., CHAMBERS, J. C., CHARLTON, R., CHATTERJEE, K., CHEN, L., CIAMPI, A., CIRAK, S., CLAPHAM, P., CLEMENT, G., COATES, G., COCCA, M., COLLIER, D. A., COSGROVE, C., COX, T., CRADDOCK, N., CROOKS, L., CURRAN, S., CURTIS, D., DALY, A., DAY, I. N. M., DAY-WILLIAMS, A., DEDOUSSIS, G., DOWN, T., DU, Y., VAN DUIJN, C. M., DUNHAM, I., EDKINS, S., EKONG, R., ELLIS, P., EVANS, D. M., FAROOQI, I. S., FITZPATRICK, D. R., FLICEK, P., FLOYD, J., FOLEY, A. R., FRANKLIN, C. S., FUTEMA, M., GALLAGHER, L., GASPARINI, P., GAUNT, T. R., GEIHS, M., GESCHWIND, D., GREENWOOD, C., GRIFFIN, H., GROZEVA, D., GUO, X., GUO, X., GURLING, H., HART, D., HENDRICKS, A. E., HOLMANS, P., HUANG, L., HUBBARD, T., HUMPHRIES, S. E., HURLES, M. E., HYSI, P., IOTCHKOVA, V., ISAACS, A., JACKSON, D. K., JAMSHIDI, Y., JOHNSON, J., JOYCE, C., KARCZEWSKI, K. J., KAYE, J., KEANE, T., KEMP, J. P., et al. 2015. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature Communications,* 6**,** 8111.

HUHN, S., DA SILVA FILHO, M. I., SANMUGANANTHAM, T., PICHULIK, T., CATALANO, C., PARDINI, B., NACCARATI, A., POLAKOVA-VYMETÁLKOVA, V., JIRASKOVA, K., VODICKOVA, L., VODICKA, P., LÖFFLER, M. W., COURTH, L., WEHKAMP, J., DIN, F. V. N., TIMOFEEVA, M., FARRINGTON, S. M., JANSEN, L., HEMMINKI, K., CHANG-CLAUDE, J., BRENNER, H., HOFFMEISTER, M., DUNLOP, M. G., WEBER, A. N. R. & FÖRSTI, A. 2018. Coding variants in NOD-like receptors: An association study on risk and survival of colorectal cancer. *PLoS One,* 13**,** e0199350.

HUISKENS, J., OLTHOF, P. B., KEIJSER, A., VAN LIENDEN, K. P., ENGELBRECHT, M. R., HERMANS, J. J., MOLENAAR, I., VERHOEF, C., DE JONG, K., DEJONG, C., KAZEMIER, G., RUERS, T., DE WILT, J., RIJKEN, A., GERHARDS, M., LIEM, M., PATIJN, G., VAN OIJEN, M., PUNT, C. & VAN GULIK, T. 2019. Assessment of colorectal liver metastasis: results of the Dutch Colorectal Cancer Group (DCCG) liver metastases expert panel of the CAIRO5 study. *HPB,* 21**,** S578.

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. & 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. & KUPFER, S. S. 2015. Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. *BMC Genomics,* 16**,** 138.

HUNG, H. Y., CHEN, J. S., CHIEN, Y., TANG, R., HSIEH, P. S., WEN, S., YOU, Y. T., YOU, J. F. & CHIANG, J. M. 2017. Preoperative alkaline phosphatase elevation was associated with poor survival in colorectal cancer patients. *Int J Colorectal Dis,* 32**,** 1775-1778.

HUTCHINS, G., SOUTHWARD, K., HANDLEY, K., MAGILL, L., BEAUMONT, C., STAHLSCHMIDT, J., RICHMAN, S., CHAMBERS, P., SEYMOUR, M., KERR, D., GRAY, R. & 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**,** 1261-70.

HUYGHE, J. R., BIEN, S. A., HARRISON, T. A., KANG, H. M., CHEN, S., SCHMIT, S. L., CONTI, D. V., QU, C., JEON, J., EDLUND, C. K., GREENSIDE, P., WAINBERG, M., SCHUMACHER, F. R., SMITH, J. D., LEVINE, D. M., NELSON, S. C., SINNOTT-ARMSTRONG, N. A., ALBANES, D., ALONSO, M. H., ANDERSON, K., ARNAU-COLLELL, C., ARNDT, V., BAMIA, C., BANBURY, B. L., BARON, J. A., BERNDT, S. I., BÉZIEAU, S., BISHOP, D. T., BOEHM, J., BOEING, H., BRENNER, H., BREZINA, S., BUCH, S., BUCHANAN, D. D., BURNETT-HARTMAN, A., BUTTERBACH, K., CAAN, B. J., CAMPBELL, P. T., CARLSON, C. S., CASTELLVÍ-BEL, S., CHAN, A. T., CHANG-CLAUDE, J., CHANOCK, S. J., CHIRLAQUE, M. D., CHO, S. H., CONNOLLY, C. M., CROSS, A. J., CUK, K., CURTIS, K. R., DE LA CHAPELLE, A., DOHENY, K. F., DUGGAN, D., EASTON, D. F., ELIAS, S. G., ELLIOTT, F., ENGLISH, D. R., FESKENS, E. J. M., FIGUEIREDO, J. C., FISCHER, R., FITZGERALD, L. M., FORMAN, D., GALA, M., GALLINGER, S., GAUDERMAN, W. J., GILES, G. G., GILLANDERS, E., GONG, J., GOODMAN, P. J., GRADY, W. M., GROVE, J. S., GSUR, A., GUNTER, M. J., HAILE, R. W., HAMPE, J., HAMPEL, H., HARLID, S., HAYES, R. B., HOFER, P., HOFFMEISTER, M., HOPPER, J. L., HSU, W. L., HUANG, W. Y., HUDSON, T. J., HUNTER, D. J., IBAÑEZ-SANZ, G., IDOS, G. E., INGERSOLL, R., JACKSON, R. D., JACOBS, E. J., JENKINS, M. A., JOSHI, A. D., JOSHU, C. E., KEKU, T. O., KEY, T. J., KIM, H. R., KOBAYASHI, E., KOLONEL, L. N., KOOPERBERG, C., KÜHN, T., KÜRY, S., et al. 2019. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet,* 51**,** 76-87.

IDOS, G. E., KWOK, J., BONTHALA, N., KYSH, L., GRUBER, S. B. & QU, C. 2020. The Prognostic Implications of Tumor Infiltrating Lymphocytes in Colorectal Cancer: A Systematic Review and Meta-Analysis. *Scientific Reports,* 10**,** 3360.

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. & 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**,** 562-578.

JAKUBOWSKA, K., KODA, M., KISIELEWSKI, W., KAŃCZUGA-KODA, L., GRUDZIŃSKA, M. & FAMULSKI, W. 2020. Pre- and postoperative neutrophil and lymphocyte count and neutrophil-to-lymphocyte ratio in patients with colorectal cancer. *Mol Clin Oncol,* 13**,** 56.

JANG, Y. S., STEINHAGEN, R. M. & HEIMANN, T. M. 1997. Colorectal cancer in familial adenomatous polyposis. *Dis Colon Rectum,* 40**,** 312-6.

JANNOT, A. S., EHRET, G. & PERNEGER, T. 2015. P < 5 × 10(-8) has emerged as a standard of statistical significance for genome-wide association studies. *J Clin Epidemiol,* 68**,** 460-5.

JEN, J., KIM, H., PIANTADOSI, S., LIU, Z. F., LEVITT, R. C., SISTONEN, P., KINZLER, K. W., VOGELSTEIN, B. & HAMILTON, S. R. 1994. Allelic loss of chromosome 18q and prognosis in colorectal cancer. *N Engl J Med,* 331**,** 213-21.

JEONG, Y. S. & BAE, Y. S. 2020. Formyl peptide receptors in the mucosal immune system. *Exp Mol Med,* 52**,** 1694-1704.

JIANG, X., KIM, H. E., SHU, H., ZHAO, Y., ZHANG, H., KOFRON, J., DONNELLY, J., BURNS, D., NG, S. C., ROSENBERG, S. & WANG, X. 2003. Distinctive roles of PHAP proteins and prothymosin-alpha in a death regulatory pathway. *Science,* 299**,** 223-6.

JIANG, Z., XU, Y. & CAI, S. 2011. Down-regulated GAS1 expression correlates with recurrence in stage II and III colorectal cancer. *Hum Pathol,* 42**,** 361-8.

JIAOJIAO, Z., HUI-ZHI, Y., SHUANG, L., MD OBAIDUL, I., YUE, Z., RONGYI, C. & ZUHUA, W. 2021. PCDH9 target RAC1/ROS-NADPH-oxidase complex with the assistance of Pyk2 interaction through Cyclin D1 trafficking. *Research Square*.

JO, W. S. & CARETHERS, J. M. 2006. Chemotherapeutic implications in microsatellite unstable colorectal cancer. *Cancer Biomark,* 2**,** 51-60.

JOHNSTON, P. G. & KAYE, S. 2001. Capecitabine: a novel agent for the treatment of solid tumors. *Anticancer Drugs,* 12**,** 639-46.

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. & GROTHEY, A. 2017. (Non-V600) BRAF Mutations Define a Clinically Distinct Molecular Subtype of Metastatic Colorectal Cancer. *J Clin Oncol,* 35**,** 2624-2630.

JOYCE, C., RAYI, A. & KASI, A. 2021. Tumor-Suppressor Genes. *StatPearls.* Treasure Island (FL): StatPearls Publishing

KALER, P., GALEA, V., AUGENLICHT, L. & KLAMPFER, L. 2010. Tumor associated macrophages protect colon cancer cells from TRAIL-induced apoptosis through IL-1beta-dependent stabilization of Snail in tumor cells. *PLoS One,* 5**,** e11700.

KAMNERDSUPAPHON, P., LORVIDHAYA, V., CHITAPANARUX, I., TONUSIN, A. & SUKTHOMYA, V. 2007. FOLFIRI chemotherapy for metastatic colorectal cancer patients. *J Med Assoc Thai,* 90**,** 2121-7.

KAMRAN, S. C., CLARK, J. W., ZHENG, H., BORGER, D. R., BLASZKOWSKY, L. S., ALLEN, J. N., KWAK, E. L., WO, J. Y., PARIKH, A. R., NIPP, R. D., MURPHY, J. E., GOYAL, L., ZHU, A. X., IAFRATE, A. J., CORCORAN, R. B., RYAN, D. P. & HONG, T. S. 2018. Primary tumor sidedness is an independent prognostic marker for survival in metastatic colorectal cancer: Results from a large retrospective cohort with mutational analysis. *Cancer Med,* 7**,** 2934-42.

KANAMARU, R., KAKUTA, H., SATO, T., ISHIOKA, C. & WAKUI, A. 1986. The inhibitory effects of 5-fluorouracil on the metabolism of preribosomal and ribosomal RNA in L-1210 cells in vitro. *Cancer Chemother Pharmacol,* 17**,** 43-6.

KAPLAN, R. 2015. The FOCUS4 design for biomarker stratified trials. *Chin Clin Oncol,* 4**,** 35.
KAPUR, K. 2017. Chapter 14 - Principles of Biostatistics. *In:* ROBERTSON, D. & WILLIAMS, G. H. (eds.) *Clinical and Translational Science (Second Edition).* Academic Press.
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. & ZALCBERG, J. R. 2008. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med,* 359**,** 1757-65.
KARSSEN, L. C., VAN DUIJN, C. M. & AULCHENKO, Y. S. 2016. The GenABEL Project for statistical genomics. *F1000Res,* 5**,** 914.
KASHAF, M. S. & MCGILL, E. 2015. Does Shared Decision Making in Cancer Treatment Improve Quality of Life? A Systematic Literature Review. *Med Decis Making,* 35**,** 1037-48.
KASSAMBARA, A., KOSINSKI, M. & BIECEK, P. 2021. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.9.
KASTRINOS, F. & SYNGAL, S. 2011. Inherited Colorectal Cancer Syndromes. *The Cancer Journal,* 17.
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. & 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**,** 370-4.
KATZ, S. J., BELKORA, J. & ELWYN, G. 2014. Shared decision making for treatment of cancer: challenges and opportunities. *J Oncol Pract,* 10**,** 206-8.
KAWAKAMI, H., ZAANAN, A. & SINICROPE, F. A. 2015. Microsatellite instability testing and its role in the management of colorectal cancer. *Curr Treat Options Oncol,* 16**,** 30.
KAWASAKI, T. & KAWAI, T. 2014. Toll-like receptor signaling pathways. *Front Immunol,* 5**,** 461.
KE, X., HUNT, S., TAPPER, W., LAWRENCE, R., STAVRIDES, G., GHORI, J., WHITTAKER, P., COLLINS, A., MORRIS, A. P., BENTLEY, D., CARDON, L. R. & DELOUKAS, P. 2004. The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum Mol Genet,* 13**,** 577-88.
KELLIS, M., WOLD, B., SNYDER, M. P., BERNSTEIN, B. E., KUNDAJE, A., MARINOV, G. K., WARD, L. D., BIRNEY, E., CRAWFORD, G. E., DEKKER, J., DUNHAM, I., ELNITSKI, L. L., FARNHAM, P. J., FEINGOLD, E. A., GERSTEIN, M., GIDDINGS, M. C., GILBERT, D. M., GINGERAS, T. R., GREEN, E. D., GUIGO, R., HUBBARD, T., KENT, J., LIEB, J. D., MYERS, R. M., PAZIN, M. J., REN, B., STAMATOYANNOPOULOS, J. A., WENG, Z., WHITE, K. P. & HARDISON, R. C. 2014. Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 6131-6138.
KENNEDY, R. & CELIS, E. 2008. Multiple roles for CD4+ T cells in anti-tumor immune responses. *Immunol Rev,* 222**,** 129-44.
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. & 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. *Lancet Oncol,* 17**,** 1543-1557.

KHATTAK, M. A., MARTIN, H., DAVIDSON, A. & PHILLIPS, M. 2015. Role of first-line anti-epidermal growth factor receptor therapy compared with anti-vascular endothelial growth factor therapy in advanced colorectal cancer: a meta-analysis of randomized clinical trials. *Clin Colorectal Cancer,* 14**,** 81-90.

KHEIRELSEID, E. A., MILLER, N., CHANG, K. H., NUGENT, M. & KERIN, M. J. 2013. Clinical applications of gene expression in colorectal cancer. *J Gastrointest Oncol,* 4**,** 144-57.

KIDWELL, K. M., YOTHERS, G., GANZ, P. A., LAND, S. R., KO, C. Y., CECCHINI, R. S., KOPEC, J. A. & WOLMARK, N. 2012. Long-term neurotoxicity effects of oxaliplatin added to fluorouracil and leucovorin as adjuvant therapy for colon cancer: results from National Surgical Adjuvant Breast and Bowel Project trials C-07 and LTS-01. *Cancer,* 118**,** 5614-22.

KIERNAN, M. C. & KRISHNAN, A. V. 2006. The pathophysiology of oxaliplatin-induced neurotoxicity. *Curr Med Chem,* 13**,** 2901-7.

KIEZUN, A., GARIMELLA, K., DO, R., STITZIEL, N. O., NEALE, B. M., MCLAREN, P. J., GUPTA, N., SKLAR, P., SULLIVAN, P. F., MORAN, J. L., HULTMAN, C. M., LICHTENSTEIN, P., MAGNUSSON, P., LEHNER, T., SHUGART, Y. Y., PRICE, A. L., DE BAKKER, P. I., PURCELL, S. M. & SUNYAEV, S. R. 2012. Exome sequencing and the genetic basis of complex traits. *Nat Genet,* 44**,** 623-30.

KIM, C. H., HUH, J. W., KIM, H. R. & KIM, Y. J. 2017. CpG island methylator phenotype is an independent predictor of survival after curative resection for colorectal cancer: A prospective cohort study. *J Gastroenterol Hepatol,* 32**,** 1469-1474.

KIM, E. R. & CHANG, D. K. 2014. Colorectal cancer in inflammatory bowel disease: the risk, pathogenesis, prevention and diagnosis. *World J Gastroenterol,* 20**,** 9872-81.

KIM, K., PANG, K. M., EVANS, M. & HAY, E. D. 2000. Overexpression of beta-catenin induces apoptosis independent of its transactivation function with LEF-1 or the involvement of major G1 cell cycle regulators. *Mol Biol Cell,* 11**,** 3509-23.

KINZLER, K. W. & VOGELSTEIN, B. 1996. Lessons from hereditary colorectal cancer. *Cell,* 87**,** 159-70.

KINZLER, K. W. & VOGELSTEIN, B. 1997. Gatekeepers and caretakers. *Nature,* 386**,** 761-763.

KLIMOSCH, S. N., FÖRSTI, A., ECKERT, J., KNEZEVIC, J., BEVIER, M., VON SCHÖNFELS, W., HEITS, N., WALTER, J., HINZ, S., LASCORZ, J., HAMPE, J., HARTL, D., FRICK, J. S., HEMMINKI, K., SCHAFMAYER, C. & WEBER, A. N. 2013. Functional TLR5 genetic variants affect human colorectal cancer survival. *Cancer Res,* 73**,** 7232-42.

KNUDSON, A. G. 1993. Antioncogenes and human cancer. *Proc Natl Acad Sci U S A,* 90**,** 10914-21.

KNUDSON, A. G. 1996. Hereditary cancer: two hits revisited. *J Cancer Res Clin Oncol,* 122**,** 135-40.

KNUDSON, A. G., JR. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A,* 68**,** 820-3.

KÖHNE, 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., SCHÖFFSKI, P., MICHEEL, S. & 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**,** 308-17.

KON, G. A. & ROSS, W. C. 1948. Effects upon tumours of various haloalkylarylamines. *Nature,* 162**,** 824.

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. & 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**,** 135-142.

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. & TABERNERO, J. 2019. Encorafenib, Binimetinib, and Cetuximab in BRAF V600E-Mutated Colorectal Cancer. *N Engl J Med,* 381**,** 1632-1643.

KORN, E. L., FREIDLIN, B. & ABRAMS, J. S. 2011. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. *J Clin Oncol,* 29**,** 2439-42.

KORN, T., BETTELLI, E., OUKKA, M. & KUCHROO, V. K. 2009. IL-17 and Th17 Cells. *Annu Rev Immunol,* 27**,** 485-517.

KOTHARI, N., SCHELL, M. J., TEER, J. K., YEATMAN, T., SHIBATA, D. & KIM, R. 2014. Comparison of KRAS mutation analysis of colorectal cancer samples by standard testing and next-generation sequencing. *J Clin Pathol,* 67**,** 764-7.

KRAKOWSKA, M., DEBSKA-SZMICH, S., CZERNEK, U., CZYZYKOWSKI, R. & POTEMSKI, P. 2016. P-032: A complete blood count as a prognostic marker in patients with metastatic colorectal cancer (mCRC). *Annals of Oncology,* 27**,** ii9-ii10.

KROEMER, G., SENOVILLA, L., GALLUZZI, L., ANDRE, F. & ZITVOGEL, L. 2015. Natural and therapy-induced immunosurveillance in breast cancer. *Nat Med,* 21**,** 1128-38.

KUIPERS, E. J., GRADY, W. M., LIEBERMAN, D., SEUFFERLEIN, T., SUNG, J. J., BOELENS, P. G., VAN DE VELDE, C. J. & WATANABE, T. 2015. Colorectal cancer. *Nat Rev Dis Primers,* 1**,** 15065.

KUMAR, S., AMBROSINI, G. & BUCHER, P. 2017. SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res,* 45**,** D139-d144.

KUMMAR, S., GUTIERREZ, M., DOROSHOW, J. H. & MURGO, A. J. 2006. Drug development in oncology: classical cytotoxics and molecularly targeted agents. *Br J Clin Pharmacol,* 62**,** 15-26.

KWEEKEL, D. M., ANTONINI, N. F., NORTIER, J. W., PUNT, C. J., GELDERBLOM, H. & GUCHELAAR, H. J. 2009. Explorative study to identify novel candidate genes related to oxaliplatin efficacy and toxicity using a DNA repair array. *Br J Cancer,* 101**,** 357-62.

LAMBERT, S. A., ABRAHAM, G. & INOUYE, M. 2019. Towards clinical utility of polygenic risk scores. *Hum Mol Genet,* 28**,** R133-r142.

LAMLUM, H., ILYAS, M., ROWAN, A., CLARK, S., JOHNSON, V., BELL, J., FRAYLING, I., EFSTATHIOU, J., PACK, K., PAYNE, S., ROYLANCE, R., GORMAN, P., SHEER, D., NEALE, K., PHILLIPS, R., TALBOT, I., BODMER, W. & TOMLINSON, I. 1999. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. *Nat Med,* 5**,** 1071-5.

LANDER, E. & KRUGLYAK, L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet,* 11**,** 241-7.

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., 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. 2019. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun,* 10**,** 2154.

LEK, M., KARCZEWSKI, K. J., MINIKEL, E. V., SAMOCHA, K. E., BANKS, E., FENNELL, T., O'DONNELL-LURIA, A. H., WARE, J. S., HILL, A. J., CUMMINGS, B. B., TUKIAINEN, T., BIRNBAUM, D. P., KOSMICKI, J. A., DUNCAN, L. E., ESTRADA, K., ZHAO, F., ZOU, J., PIERCE-HOFFMAN, E., BERGHOUT, J., COOPER, D. N., DEFLAUX, N., DEPRISTO, M., DO, R., FLANNICK, J., FROMER, M., GAUTHIER, L., GOLDSTEIN, J., GUPTA, N., HOWRIGAN, D., KIEZUN, A., KURKI, M. I., MOONSHINE, A. L., NATARAJAN, P., OROZCO, L., PELOSO, G. M., POPLIN, R., RIVAS, M. A., RUANO-RUBIO, V., ROSE, S. A., RUDERFER, D. M., SHAKIR, K., STENSON, P. D., STEVENS, C., THOMAS, B. P., TIAO, G., TUSIE-LUNA,

M. T., WEISBURD, B., WON, H. H., YU, D., ALTSHULER, D. M., ARDISSINO, D., BOEHNKE, M., DANESH, J., DONNELLY, S., ELOSUA, R., FLOREZ, J. C., GABRIEL, S. B., GETZ, G., GLATT, S. J., HULTMAN, C. M., KATHIRESAN, S., LAAKSO, M., MCCARROLL, S., MCCARTHY, M. I., MCGOVERN, D., MCPHERSON, R., NEALE, B. M., PALOTIE, A., PURCELL, S. M., SALEHEEN, D., SCHARF, J. M., SKLAR, P., SULLIVAN, P. F., TUOMILEHTO, J., TSUANG, M. T., WATKINS, H. C., WILSON, J. G., DALY, M. J. & MACARTHUR, D. G. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature,* 536**,** 285-91.

LEOZ, M. L., CARBALLAL, S., MOREIRA, L., OCAÑA, T. & BALAGUER, F. 2015. The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management. *Appl Clin Genet,* 8**,** 95-107.

LESLIE, A., CAREY, F. A., PRATT, N. R. & STEELE, R. J. C. 2002. The colorectal adenoma–carcinoma sequence. *British Journal of Surgery,* 89**,** 845-860.

LEVITT, N. C. & HICKSON, I. D. 2002. Caretaker tumour suppressor genes that defend genome integrity. *Trends Mol Med,* 8**,** 179-86.

LEWIS, C. M. & VASSOS, E. 2020. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine,* 12**,** 44.

LEWONTIN, R. C. 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics,* 49**,** 49-67.

LI, L., CHEN, K., XIANG, Y., YOSHIMURA, T., SU, S., ZHU, J., BIAN, X.-W. & WANG, J. M. 2016. New development in studies of formyl-peptide receptors: critical roles in host defense. *Journal of leukocyte biology,* 99**,** 425-435.

LI, L., FAN, P., CHOU, H., LI, J., WANG, K. & LI, H. 2019a. Herbacetin suppressed MMP9 mediated angiogenesis of malignant melanoma through blocking EGFR-ERK/AKT signaling pathway. *Biochimie,* 162**,** 198-207.

LI, M., LI, C. & GUAN, W. 2008. Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur J Hum Genet,* 16**,** 635-43.

LI, R., LIANG, M., LIANG, X., YANG, L., SU, M. & LAI, K. P. 2020. Chemotherapeutic Effectiveness of Combining Cetuximab for Metastatic Colorectal Cancer Treatment: A System Review and Meta-Analysis. *Front Oncol,* 10**,** 868.

LI, S., XU, K., GU, D., HE, L., XIE, L., CHEN, Z., FAN, Z., ZHU, L., DU, M., CHU, H., ZHANG, Z., WU, Y., NI, M. & WANG, M. 2019b. Genetic variants in RPA1 associated with the response to oxaliplatin-based chemotherapy in colorectal cancer. *J Gastroenterol,* 54**,** 939-949.

LI, S.-Q., SU, N., GONG, P., ZHANG, H.-B., LIU, J., WANG, D., SUN, Y.-P., ZHANG, Y., QIAN, F., ZHAO, B., YU, Y. & YE, R. D. 2017. The Expression of Formyl Peptide Receptor 1 is Correlated with Tumor Invasion of Human Colorectal Cancer. *Scientific Reports,* 7**,** 5918.

LI, W., QIU, T., ZHI, W., SHI, S., ZOU, S., LING, Y., SHAN, L., YING, J. & LU, N. 2015. Colorectal carcinomas with KRAS codon 12 mutation are associated with more advanced tumor stages. *BMC cancer,* 15**,** 340-340.

LI, Y., WILLER, C., SANNA, S. & ABECASIS, G. 2009. Genotype imputation. *Annu Rev Genomics Hum Genet,* 10**,** 387-406.

LIANG, Y. H., LIN, Y. L., LIAU, J. Y., TSAI, J. H., LIANG, J. T., LIN, B. R., HUNG, J. S., TSENG, L. H., LIN, L. I., CHANG, Y. L., CHENG, A. L. & YEH, K. H. 2015. Cetuximab Might Be Detrimental to Metastatic Colorectal Cancer Patients with KRAS Codon 12 Mutations. *Anticancer Res,* 35**,** 4207-14.

LIÈVRE, A., BACHET, J. B., LE CORRE, D., BOIGE, V., LANDI, B., EMILE, J. F., CÔTÉ, J. F., TOMASIC, G., PENNA, C., DUCREUX, M., ROUGIER, P., PENAULT-LLORCA, F. & LAURENT-PUIG, P. 2006. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res,* 66**,** 3992-5.

LIN, Y. L., LIANG, Y. H., TSAI, J. H., LIAU, J. Y., LIANG, J. T., LIN, B. R., HUNG, J. S., LIN, L. I., TSENG, L. H., CHANG, Y. L., YEH, K. H. & CHENG, A. L. 2014. Oxaliplatin-based chemotherapy is more beneficial in KRAS mutant than in KRAS wild-type metastatic colorectal cancer patients. *PLoS One,* 9**,** e86789.

LIU, C.-J., FU, X., XIA, M., ZHANG, Q., GU, Z. & GUO, A.-Y. 2020. miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. *Nucleic Acids Research,* 49**,** D1276-D1281.

LIU, M., ZHAO, J., CHEN, K., BIAN, X., WANG, C., SHI, Y. & WANG, J. M. 2012. G protein-coupled receptor FPR1 as a pharmacologic target in inflammation and human glioblastoma. *Int Immunopharmacol,* 14**,** 283-8.

LIU, R., TRUAX, A. D., CHEN, L., HU, P., LI, Z., CHEN, J., SONG, C., CHEN, L. & TING, J. P. 2015. Expression profile of innate immune receptors, NLRs and AIM2, in human colorectal cancer: correlation with cancer stages and inflammasome components. *Oncotarget,* 6**,** 33456-69.

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. & OGINO, S. 2013. Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. *J Natl Cancer Inst,* 105**,** 1151-6.

LOGAN, C. Y. & NUSSE, R. 2004. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol,* 20**,** 781-810.

LONG, H., O'CONNOR, B. P., ZEMANS, R. L., ZHOU, X., YANG, I. V. & SCHWARTZ, D. A. 2014. The Toll-like receptor 4 polymorphism Asp299Gly but not Thr399Ile influences TLR4 signaling and function. *PLoS One,* 9**,** e93550.

LONGLEY, D. B., HARKIN, D. P. & JOHNSTON, P. G. 2003. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer,* 3**,** 330-338.

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. & LE MARCHAND, L. 2012. cis-Expression QTL analysis of established colorectal cancer risk variants in colon tumors and adjacent normal tissue. *PLoS One,* 7**,** e30477.

LOWE, E. L., CROTHER, T. R., RABIZADEH, S., HU, B., WANG, H., CHEN, S., SHIMADA, K., WONG, M. H., MICHELSEN, K. S. & ARDITI, M. 2010. Toll-like receptor 2 signaling protects mice from tumor development in a mouse model of colitis-induced cancer. *PLoS One,* 5**,** e13027.

LU, J., ZHAO, J., JIA, C., ZHOU, L., CAI, Y., NI, J., MA, J., ZHENG, M. & LU, A. 2019. FPR2 enhances colorectal cancer progression by promoting EMT process. *Neoplasma,* 66**,** 785-791.

LV, W., ZHANG, G. Q., JIAO, A., ZHAO, B. C., SHI, Y., CHEN, B. M. & ZHANG, J. L. 2017. Chemotherapy Plus Cetuximab versus Chemotherapy Alone for Patients with KRAS Wild Type Unresectable Liver-Confined Metastases

Colorectal Cancer: An Updated Meta-Analysis of RCTs. *Gastroenterol Res Pract,* 2017**,** 8464905.

LYNCH, H. T. & DE LA CHAPELLE, A. 2003. Hereditary colorectal cancer. *N Engl J Med,* 348**,** 919-32.

MACARTHUR, D. G., MANOLIO, T. A., DIMMOCK, D. P., REHM, H. L., SHENDURE, J., ABECASIS, G. R., ADAMS, D. R., ALTMAN, R. B., ANTONARAKIS, S. E., ASHLEY, E. A., BARRETT, J. C., BIESECKER, L. G., CONRAD, D. F., COOPER, G. M., COX, N. J., DALY, M. J., GERSTEIN, M. B., GOLDSTEIN, D. B., HIRSCHHORN, J. N., LEAL, S. M., PENNACCHIO, L. A., STAMATOYANNOPOULOS, J. A., SUNYAEV, S. R., VALLE, D., VOIGHT, B. F., WINCKLER, W. & GUNTER, C. 2014. Guidelines for investigating causality of sequence variants in human disease. *Nature,* 508**,** 469-76.

MACHIELA, M. J. & CHANOCK, S. J. 2015. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics,* 31**,** 3555-7.

MACHOVER, D., DIAZ-RUBIO, E., DE GRAMONT, A., SCHILF, A., GASTIABURU, J. J., BRIENZA, S., ITZHAKI, M., METZGER, G., N'DAW, D., VIGNOUD, J., ABAD, A., FRANCOIS, E., GAMELIN, E., MARTY, M., SASTRE, J., SEITZ, J. F. & YCHOU, M. 1996. Two consecutive phase II studies of oxaliplatin (L-OHP) for treatment of patients with advanced colorectal carcinoma who were resistant to previous treatment with fluoropyrimidines. *Ann Oncol,* 7**,** 95-8.

MACKENZIE, A., HING, B. & DAVIDSON, S. 2013. Exploring the effects of polymorphisms on cis-regulatory signal transduction response. *Trends Mol Med,* 19**,** 99-107.

MACLEOD, M. R. 2002. PubMed: http://www.pubmed.org. *J Neurol Neurosurg Psychiatry,* 73**,** 746.

MAJEK, O., GONDOS, A., JANSEN, L., EMRICH, K., HOLLECZEK, B., KATALINIC, A., NENNECKE, A., EBERLE, A. & BRENNER, H. 2013. Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in Germany. *PLoS One,* 8**,** e68077.

MALMBERG, K. J., BRYCESON, Y. T., CARLSTEN, M., ANDERSSON, S., BJORKLUND, A., BJORKSTROM, N. K., BAUMANN, B. C., FAURIAT, C., ALICI, E., DILBER, M. S. & LJUNGGREN, H. G. 2008. NK cell-mediated targeting of human cancer and possibilities for new means of immunotherapy. *Cancer Immunol Immunother,* 57**,** 1541-52.

MAO, C., HUANG, Y. F., YANG, Z. Y., ZHENG, D. Y., CHEN, J. Z. & TANG, J. L. 2013. KRAS p.G13D mutation and codon 12 mutations are not created equal in predicting clinical outcomes of cetuximab in metastatic colorectal cancer: a systematic review and meta-analysis. *Cancer,* 119**,** 714-21.

MAO, Z., ZHAO, H., QIN, Y., WEI, J., SUN, J., ZHANG, W. & KANG, Y. 2020. Post-Transcriptional Dysregulation of microRNA and Alternative Polyadenylation in Colorectal Cancer. *Front Genet,* 11**,** 64.

MARCHINI, J. & HOWIE, B. 2010. Genotype imputation for genome-wide association studies. *Nat Rev Genet,* 11**,** 499-511.

MARCHINI, J., HOWIE, B., MYERS, S., MCVEAN, G. & DONNELLY, P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet,* 39**,** 906-13.

MARKMAN, J. L. & SHIAO, S. L. 2015. Impact of the immune system and immunotherapy in colorectal cancer. *J Gastrointest Oncol,* 6**,** 208-23.

MÁRMOL, I., SÁNCHEZ-DE-DIEGO, C., PRADILLA DIESTE, A., CERRADA, E. & RODRIGUEZ YOLDI, M. J. 2017. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci,* 18.

MAROULI, E., GRAFF, M., MEDINA-GOMEZ, C., LO, K. S., WOOD, A. R., KJAER, T. R., FINE, R. S., LU, Y., SCHURMANN, C., HIGHLAND, H. M., RÜEGER, S., THORLEIFSSON, G., JUSTICE, A. E., LAMPARTER, D., STIRRUPS, K. E., TURCOT, V., YOUNG, K. L., WINKLER, T. W., ESKO, T., KARADERI, T., LOCKE, A. E., MASCA, N. G., NG, M. C., MUDGAL, P., RIVAS, M. A., VEDANTAM, S., MAHAJAN, A., GUO, X., ABECASIS, G., ABEN, K. K., ADAIR, L. S., ALAM, D. S., ALBRECHT, E., ALLIN, K. H., ALLISON, M., AMOUYEL, P., APPEL, E. V., ARVEILER, D., ASSELBERGS, F. W., AUER, P. L., BALKAU, B., BANAS, B., BANG, L. E., BENN, M., BERGMANN, S., BIELAK, L. F., BLÜHER, M., BOEING, H., BOERWINKLE, E., BÖGER, C. A., BONNYCASTLE, L. L., BORK-JENSEN, J., BOTS, M. L., BOTTINGER, E. P., BOWDEN, D. W., BRANDSLUND, I., BREEN, G., BRILLIANT, M. H., BROER, L., BURT, A. A., BUTTERWORTH, A. S., CAREY, D. J., CAULFIELD, M. J., CHAMBERS, J. C., CHASMAN, D. I., CHEN, Y. I., CHOWDHURY, R., CHRISTENSEN, C., CHU, A. Y., COCCA, M., COLLINS, F. S., COOK, J. P., CORLEY, J., GALBANY, J. C., COX, A. J., CUELLAR-PARTIDA, G., DANESH, J., DAVIES, G., DE BAKKER, P. I., DE BORST, G. J., DE DENUS, S., DE GROOT, M. C., DE MUTSERT, R., DEARY, I. J., DEDOUSSIS, G., DEMERATH, E. W., DEN HOLLANDER, A. I., DENNIS, J. G., DI ANGELANTONIO, E., DRENOS, F., DU, M., DUNNING, A. M., EASTON, D. F., EBELING, T., EDWARDS, T. L., ELLINOR, P. T., ELLIOTT, P., EVANGELOU, E., FARMAKI, A. E., FAUL, J. D., et al. 2017. Rare and low-frequency coding variants alter human adult height. *Nature,* 542**,** 186-190.

MARSHALL, J. S., WARRINGTON, R., WATSON, W. & KIM, H. L. 2018. An introduction to immunology and immunopathology. *Allergy Asthma Clin Immunol,* 14**,** 49.

MASSY, W. F. 1965. Principal Components Regression in Exploratory Statistical Research. *Journal of the American Statistical Association,* 60**,** 234-256.

MATIĆ, I. Z., KOLUNDŽIJA, B., DAMJANOVIĆ, A., SPASIĆ, J., RADOSAVLJEVIĆ, D., ĐORĐIĆ CRNOGORAC, M., GROZDANIĆ, N. & JURANIĆ, Z. D. 2017. Peripheral White Blood Cell Subsets in Metastatic Colorectal Cancer Patients Treated with Cetuximab: The Potential Clinical Relevance. *Front Immunol,* 8**,** 1886.

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. & 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**,** 2103-14.

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. & 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**,** 2178-86.

MAURANO, M. T., HUMBERT, R., RYNES, E., THURMAN, R. E., HAUGEN, E., WANG, H., REYNOLDS, A. P., SANDSTROM, R., QU, H., BRODY, J., SHAFER, A., NERI, F., LEE, K., KUTYAVIN, T., STEHLING-SUN, S., JOHNSON, A. K., CANFIELD, T. K., GISTE, E., DIEGEL, M., BATES, D., HANSEN, R. S., NEPH, S., SABO, P. J., HEIMFELD, S., RAUBITSCHEK, A., ZIEGLER, S., COTSAPAS, C., SOTOODEHNIA, N., GLASS, I., SUNYAEV, S. R., KAUL, R. & STAMATOYANNOPOULOS, J. A. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science,* 337**,** 1190-5.

MAYER, R. J., VAN CUTSEM, E., FALCONE, A., YOSHINO, T., GARCIA-CARBONERO, R., MIZUNUMA, N., YAMAZAKI, K., SHIMADA, Y., TABERNERO, J., KOMATSU, Y., SOBRERO, A., BOUCHER, E., PEETERS, M., TRAN, B., LENZ, H.-J., ZANIBONI, A., HOCHSTER, H., CLEARY, J. M., PRENEN, H., BENEDETTI, F., MIZUGUCHI, H., MAKRIS, L., ITO, M. & OHTSU, A. 2015. Randomized Trial of TAS-102 for Refractory Metastatic Colorectal Cancer. *New England Journal of Medicine,* 372**,** 1909-1919.

MCCARTHY, M. I., ABECASIS, G. R., CARDON, L. R., GOLDSTEIN, D. B., LITTLE, J., IOANNIDIS, J. P. & HIRSCHHORN, J. N. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet,* 9**,** 356-69.

MCNABB, S., HARRISON, T. A., ALBANES, D., BERNDT, S. I., BRENNER, H., CAAN, B. J., CAMPBELL, P. T., CAO, Y., CHANG-CLAUDE, J., CHAN, A., CHEN, Z., ENGLISH, D. R., GILES, G. G., GIOVANNUCCI, E. L., GOODMAN, P. J., HAYES, R. B., HOFFMEISTER, M., JACOBS, E. J., JOSHI, A. D., LARSSON, S. C., LE MARCHAND, L., LI, L., LIN, Y., MÄNNISTÖ, S., MILNE, R. L., NAN, H., NEWTON, C. C., OGINO, S., PARFREY, P. S., PETERSEN, P. S., POTTER, J. D., SCHOEN, R. E., SLATTERY, M. L., SU, Y. R., TANGEN, C. M., TUCKER, T. C., WEINSTEIN, S. J., WHITE, E., WOLK, A., WOODS, M. O., PHIPPS, A. I. & PETERS, U. 2020. Meta-analysis of 16 studies of the association of alcohol with colorectal cancer. *Int J Cancer,* 146**,** 861-873.

MCSHANE, L. M., ALTMAN, D. G., SAUERBREI, W., TAUBE, S. E., GION, M. & CLARK, G. M. 2005. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer,* 93**,** 387-91.

MEI, S., QIN, Q., WU, Q., SUN, H., ZHENG, R., ZANG, C., ZHU, M., WU, J., SHI, X., TAING, L., LIU, T., BROWN, M., MEYER, C. A. & LIU, X. S. 2017. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res,* 45**,** D658-d662.

MEI, Z. B., DUAN, C. Y., LI, C. B., CUI, L. & OGINO, S. 2016. Prognostic role of tumor PIK3CA mutation in colorectal cancer: a systematic review and meta-analysis. *Ann Oncol,* 27**,** 1836-48.

MENDELAAR, P. A. J., SMID, M., VAN RIET, J., ANGUS, L., LABOTS, M., STEEGHS, N., HENDRIKS, M. P., CIRKEL, G. A., VAN ROOIJEN, J. M.,

TEN TIJE, A. J., LOLKEMA, M. P., CUPPEN, E., SLEIJFER, S., MARTENS, J. W. M. & WILTING, S. M. 2021. Whole genome sequencing of metastatic colorectal cancer reveals prior treatment effects and specific metastasis features. *Nature Communications,* 12**,** 574.

MICHOR, F., IWASA, Y. & NOWAK, M. A. 2004. Dynamics of cancer progression. *Nature Reviews Cancer,* 4**,** 197-205.

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. & 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**,** 4575-80.

MIELE, A. & DEKKER, J. 2009. Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). *Methods in molecular biology (Clifton, N.J.),* 464**,** 105-121.

MIGEOTTE, I., COMMUNI, D. & PARMENTIER, M. 2006. Formyl peptide receptors: a promiscuous subfamily of G protein-coupled receptors controlling immune responses. *Cytokine Growth Factor Rev,* 17**,** 501-19.

MILLAN, M., MERINO, S., CARO, A., FELIU, F., ESCUDER, J. & FRANCESCH, T. 2015. Treatment of colorectal cancer in the elderly. *World J Gastrointest Oncol,* 7**,** 204-20.

MILLER, K. D., NOGUEIRA, L., MARIOTTO, A. B., ROWLAND, J. H., YABROFF, K. R., ALFANO, C. M., JEMAL, A., KRAMER, J. L. & SIEGEL, R. L. 2019. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin,* 69**,** 363-385.

MITTRA, I. 2007. The disconnection between tumor response and survival. *Nature Clinical Practice Oncology,* 4**,** 203-203.

MIYOSHI, Y., NAGASE, H., ANDO, H., HORII, A., ICHII, S., NAKATSURU, S., AOKI, T., MIKI, Y., MORI, T. & NAKAMURA, Y. 1992. Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. *Hum Mol Genet,* 1**,** 229-33.

MOORE, C. & JACOBSON, S. 2020. genpwr: Power Calculations Under Genetic Model Misspecification. 1.0.2 ed.

MOORE, J. E., PURCARO, M. J., PRATT, H. E., EPSTEIN, C. B., SHORESH, N., ADRIAN, J., KAWLI, T., DAVIS, C. A., DOBIN, A., KAUL, R., HALOW, J., VAN NOSTRAND, E. L., FREESE, P., GORKIN, D. U., SHEN, Y., HE, Y., MACKIEWICZ, M., PAULI-BEHN, F., WILLIAMS, B. A., MORTAZAVI, A., KELLER, C. A., ZHANG, X.-O., ELHAJJAJY, S. I., HUEY, J., DICKEL, D. E., SNETKOVA, V., WEI, X., WANG, X., RIVERA-MULIA, J. C., ROZOWSKY, J., ZHANG, J., CHHETRI, S. B., ZHANG, J., VICTORSEN, A., WHITE, K. P., VISEL, A., YEO, G. W., BURGE, C. B., LÉCUYER, E., GILBERT, D. M., DEKKER, J., RINN, J., MENDENHALL, E. M., ECKER, J. R., KELLIS, M., KLEIN, R. J., NOBLE, W. S., KUNDAJE, A., GUIGÓ, R., FARNHAM, P. J., CHERRY, J. M., MYERS, R. M., REN, B., GRAVELEY, B. R., GERSTEIN, M. B., PENNACCHIO, L. A., SNYDER, M. P., BERNSTEIN, B. E., WOLD, B., HARDISON, R. C., GINGERAS, T. R., STAMATOYANNOPOULOS, J. A., WENG, Z. & THE, E. P. C. 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature,* 583**,** 699-710.

MORENO, V., ALONSO, M. H., CLOSA, A., VALLÉS, X., DIEZ-VILLANUEVA, A., VALLE, L., CASTELLVÍ-BEL, S., SANZ-PAMPLONA, R., LOPEZ-DORIGA,

A., CORDERO, D. & SOLÉ, X. 2018. Colon-specific eQTL analysis to inform on functional SNPs. *British Journal of Cancer,* 119**,** 971-977.
MORGAN, M. 2019. BiocManager: Access the Bioconductor Project Package Repository. R package version 1.30.10.
MORGEN, E. K., LENZ, H. J., JONKER, D. J., TU, D., MILANO, G., GRAZIANO, F., ZALCBERG, J., KARAPETIS, C. S., DOBROVIC, A., O'CALLAGHAN, C. J. & LIU, G. 2017. Germline polymorphisms as biomarkers of tumor response in colorectal cancer patients treated with anti-EGFR monoclonal antibodies: a systematic review and meta-analysis. *Pharmacogenomics J,* 17**,** 535-542.
MORTON, D. 2019. FOxTROT: An international randomised controlled trial in 1053 patients evaluating neoadjuvant chemotherapy (NAC) for colon cancer. On behalf of the FOxTROT Collaborative Group. *Annals of Oncology,* 30**,** v198.
MUELLER, M. M. & FUSENIG, N. E. 2004. Friends or foes - bipolar effects of the tumour stroma in cancer. *Nat Rev Cancer,* 4**,** 839-49.
MULLEN, A. C., ORLANDO, D. A., NEWMAN, J. J., LOVÉN, J., KUMAR, R. M., BILODEAU, S., REDDY, J., GUENTHER, M. G., DEKOTER, R. P. & YOUNG, R. A. 2011. Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell,* 147**,** 565-76.
MUNHOZ, R. R. & POSTOW, M. A. 2016. Recent advances in understanding antitumor immunity. *F1000Res,* 5**,** 2545.
MUNTEANU, I. & MASTALIER, B. 2014. Genetics of colorectal cancer. *J Med Life,* 7**,** 507-11.
MUSUNURU, K., STRONG, A., FRANK-KAMENETSKY, M., LEE, N. E., AHFELDT, T., SACHS, K. V., LI, X., LI, H., KUPERWASSER, N., RUDA, V. M., PIRRUCCELLO, J. P., MUCHMORE, B., PROKUNINA-OLSSON, L., HALL, J. L., SCHADT, E. E., MORALES, C. R., LUND-KATZ, S., PHILLIPS, M. C., WONG, J., CANTLEY, W., RACIE, T., EJEBE, K. G., ORHO-MELANDER, M., MELANDER, O., KOTELIANSKY, V., FITZGERALD, K., KRAUSS, R. M., COWAN, C. A., KATHIRESAN, S. & RADER, D. J. 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature,* 466**,** 714-719.
NARUM, S. R. 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics,* 7**,** 783-787.
NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2011. Colorectal cancer: diagnosis and management. *Clinical guideline [CG131].* NICE.
NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2017. Cetuximab and panitumumab for previously untreated metastatic colorectal cancer. *Technology appraisal guidance [TA439].* NICE.
NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2020a. Colorectal Cancer. *NICE guideline [NG151]* NICE.
NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2020b. COVID-19 rapid guideline: delivery of systemic anticancer treatments. *NICE guideline [NG161].* NICE.
NAYAK, M. G., GEORGE, A., VIDYASAGAR, M. S., MATHEW, S., NAYAK, S., NAYAK, B. S., SHASHIDHARA, Y. N. & KAMATH, A. 2017. Quality of Life among Cancer Patients. *Indian J Palliat Care,* 23**,** 445-450.

NAZEMALHOSSEINI MOJARAD, E., KUPPEN, P. J., AGHDAEI, H. A. & ZALI, M. R. 2013. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench,* 6**,** 120-8.

NEGRINI, S., GORGOULIS, V. G. & HALAZONETIS, T. D. 2010. Genomic instability — an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology,* 11**,** 220-228.

NETEA, M. G., VAN DE VEERDONK, F. L. & VAN DER MEER, J. W. 2012. Primary immunodeficiencies of pattern recognition receptors. *J Intern Med,* 272**,** 517-27.

NETEA, M. G. & VAN DER MEER, J. W. 2011. Immunodeficiency and genetic defects of pattern-recognition receptors. *N Engl J Med,* 364**,** 60-70.

NEUFELD, K. L. & WHITE, R. L. 1997. Nuclear and cytoplasmic localizations of the adenomatous polyposis coli protein. *Proc Natl Acad Sci U S A,* 94**,** 3034-9.

NEUWIRTH, E. 2014. RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.

NHS NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE (NICE) 2017. Cetuximab and panitumumab for previously untreated metastatic colorectal cancer. NICE Technology appraisal guidance.

NICA, A. C. & DERMITZAKIS, E. T. 2013. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20120362.

NICCOLAI, E., RICCI, F., RUSSO, E., NANNINI, G., EMMI, G., TADDEI, A., RINGRESSI, M. N., MELLI, F., MILOEVA, M., CIANCHI, F., BECHI, P., PRISCO, D. & AMEDEI, A. 2017. The Different Functional Distribution of "Not Effector" T Cells (Treg/Tnull) in Colorectal Cancer. *Front Immunol,* 8**,** 1900.

NIELSEN, H. J., HANSEN, U., CHRISTENSEN, I. J., REIMERT, C. M., BRÜNNER, N. & MOESGAARD, F. 1999. Independent prognostic value of eosinophil and mast cell infiltration in colorectal cancer tissue. *J Pathol,* 189**,** 487-95.

NISHIZAKI, S. S., NG, N., DONG, S., PORTER, R. S., MORTERUD, C., WILLIAMS, C., ASMAN, C., SWITZENBERG, J. A. & BOYLE, A. P. 2020. Predicting the effects of SNPs on transcription factor binding affinity. *Bioinformatics,* 36**,** 364-372.

NOJADEH, J. N., BEHROUZ SHARIF, S. & SAKHINIA, E. 2018. Microsatellite instability in colorectal cancer. *Excli j,* 17**,** 159-168.

NOSHO, K., BABA, Y., TANAKA, N., SHIMA, K., HAYASHI, M., MEYERHARDT, J. A., GIOVANNUCCI, E., DRANOFF, G., FUCHS, C. S. & OGINO, S. 2010. Tumour-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. *J Pathol,* 222**,** 350-66.

O'DWYER, P. J., MANOLA, J., VALONE, F. H., RYAN, L. M., HINES, J. D., WADLER, S., HALLER, D. G., ARBUCK, S. G., WEINER, L. M., MAYER, R. J. & BENSON, A. B., 3RD 2001. Fluorouracil modulation in colorectal cancer: lack of improvement with N -phosphonoacetyl- l -aspartic acid or oral leucovorin or interferon, but enhanced therapeutic index with weekly 24-hour infusion schedule--an Eastern Cooperative Oncology Group/Cancer and Leukemia Group B Study. *J Clin Oncol,* 19**,** 2413-21.

OBEID, M., TESNIERE, A., GHIRINGHELLI, F., FIMIA, G. M., APETOH, L., PERFETTINI, J. L., CASTEDO, M., MIGNOT, G., PANARETAKIS, T.,

CASARES, N., METIVIER, D., LAROCHETTE, N., VAN ENDERT, P., CICCOSANTI, F., PIACENTINI, M., ZITVOGEL, L. & KROEMER, G. 2007. Calreticulin exposure dictates the immunogenicity of cancer cell death. *Nat Med,* 13**,** 54-61.

ODA, K., MATSUOKA, Y., FUNAHASHI, A. & KITANO, H. 2005. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol,* 1**,** 2005.0010.

OETTING, W. S., JACOBSON, P. A. & ISRANI, A. K. 2017. Validation Is Critical for Genome-Wide Association Study-Based Associations. *Am J Transplant,* 17**,** 318-319.

OFFICE FOR NATIONAL STATISTICS. 2019. *Cancer survival in England - adults diagnosed (2013-2017)* [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalratescancersurvivalinenglandadultsdiagnosed [Accessed 1st May 2021].

OFFICE FOR NATIONAL STATISTICS. 2020. *Number of deaths from colorectal cancer, England, deaths registered between 2018 and 2019* [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/12244numberofdeathsfromcolorectalcancerenglanddeathsregisteredbetween2018and2019 [Accessed 1st May 2021].

OGINO, S., NOSHO, K., KIRKNER, G. J., KAWASAKI, T., MEYERHARDT, J. A., LODA, M., GIOVANNUCCI, E. L. & FUCHS, C. S. 2009a. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut,* 58**,** 90-6.

OGINO, S., NOSHO, K., KIRKNER, G. J., SHIMA, K., IRAHARA, N., KURE, S., CHAN, A. T., ENGELMAN, J. A., KRAFT, P., CANTLEY, L. C., GIOVANNUCCI, E. L. & FUCHS, C. S. 2009b. PIK3CA mutation is associated with poor prognosis among patients with curatively resected colon cancer. *J Clin Oncol,* 27**,** 1477-84.

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. & MOLEY, J. F. 1998. Confirmation that chromosome 18q allelic loss in colon cancer is a prognostic indicator. *J Clin Oncol,* 16**,** 427-33.

OH, J. H., KERNS, S., OSTRER, H., POWELL, S. N., ROSENSTEIN, B. & DEASY, J. O. 2017. Computational methods using genome-wide association studies to predict radiotherapy complications and to identify correlative molecular processes. *Scientific Reports,* 7**,** 43381.

ORSETTI, B., SELVES, J., BASCOUL-MOLLEVI, C., LASORSA, L., GORDIEN, K., BIBEAU, F., MASSEMIN, B., PARAF, F., SOUBEYRAN, I., HOSTEIN, I., DAPREMONT, V., GUIMBAUD, R., CAZAUX, C., LONGY, M. & THEILLET, C. 2014. Impact of chromosomal instability on colorectal cancer progression and outcome. *BMC Cancer,* 14**,** 121.

OUCHI, K., TAKAHASHI, S., YAMADA, Y., TSUJI, S., TATSUNO, K., TAKAHASHI, H., TAKAHASHI, N., TAKAHASHI, M., SHIMODAIRA, H., ABURATANI, H. & ISHIOKA, C. 2015. DNA methylation status as a biomarker of anti-epidermal growth factor receptor treatment for metastatic colorectal cancer. *Cancer Sci,* 106**,** 1722-9.

OWZAR, K., LI, Z., COX, N. & JUNG, S. H. 2012. Power and sample size calculations for SNP association studies with censored time-to-event outcomes. *Genet Epidemiol,* 36**,** 538-48.

OZCAN, M., JANIKOVITS, J., VON KNEBEL DOEBERITZ, M. & KLOOR, M. 2018. Complex pattern of immune evasion in MSI colorectal cancer. *Oncoimmunology,* 7**,** e1445453.

PADMA, V. V. 2015. An overview of targeted cancer therapy. *Biomedicine (Taipei),* 5**,** 19.

PAGES, F., BERGER, A., CAMUS, M., SANCHEZ-CABO, F., COSTES, A., MOLIDOR, R., MLECNIK, B., KIRILOVSKY, A., NILSSON, M., DAMOTTE, D., MEATCHI, T., BRUNEVAL, P., CUGNENC, P. H., TRAJANOSKI, Z., FRIDMAN, W. H. & GALON, J. 2005. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med,* 353**,** 2654-66.

PARIKH, A. R., LEE, F. C., YAU, L., KOH, H., KNOST, J., MITCHELL, E. P., BOSANAC, I., CHOONG, N., SCAPPATICCI, F., MANCAO, C. & LENZ, H. J. 2019. MAVERICC, a Randomized, Biomarker-stratified, Phase II Study of mFOLFOX6-Bevacizumab versus FOLFIRI-Bevacizumab as First-line Chemotherapy in Metastatic Colorectal Cancer. *Clin Cancer Res,* 25**,** 2988-2995.

PARK, D. J., STOEHLMACHER, J., ZHANG, W., TSAO-WEI, D. D., GROSHEN, S. & LENZ, H. J. 2001. A Xeroderma pigmentosum group D gene polymorphism predicts clinical outcome to platinum-based chemotherapy in patients with advanced colorectal cancer. *Cancer Res,* 61**,** 8654-8.

PARK, D. J., ZHANG, W., STOEHLMACHER, J., TSAO-WEI, D., GROSHEN, S., GIL, J., YUN, J., SONES, E., MALLIK, N. & LENZ, H. J. 2003. ERCC1 gene polymorphism as a predictor for clinical outcome in advanced colorectal cancer patients treated with platinum-based chemotherapy. *Clin Adv Hematol Oncol,* 1**,** 162-6.

PARSA, N. 2012. Environmental factors inducing human cancers. *Iran J Public Health,* 41**,** 1-9.

PASCHE, B. & YI, N. 2010. Candidate gene association studies: successes and failures. *Curr Opin Genet Dev,* 20**,** 257-61.

PASCHEN, A., EICHMULLER, S. & SCHADENDORF, D. 2004. Identification of tumor antigens and T-cell epitopes, and its clinical application. *Cancer Immunol Immunother,* 53**,** 196-203.

PASETTO, L. M., JIRILLO, A., IADICICCO, G., ROSSI, E., PARIS, M. K. & MONFARDINI, S. 2005. FOLFOX versus FOLFIRI: a comparison of regimens in the treatment of colorectal cancer metastases. *Anticancer Res,* 25**,** 563-76.

PENDLEBURY, S., DUCHESNE, F., REED, K. A., SMITH, J. L. & KERR, D. J. 2003. A Trial of Adjuvant Therapy in Colorectal Cancer: The VICTOR Trial. *Clinical Colorectal Cancer,* 3**,** 58-60.

PENNEY, K. L., BANBURY, B. L., BIEN, S., HARRISON, T. A., HUA, X., PHIPPS, A. I., SUN, W., SONG, M., JOSHI, A. D., ALBERTS, S. R., ALLEGRA, C. J., ATKINS, J., COLANGELO, L. H., GEORGE, T. J., GOLDBERG, R. M., LUCAS, P. C., NAIR, S. G., SHI, Q., SINICROPE, F. A., WOLMARK, N., YOTHERS, G., PETERS, U., NEWCOMB, P. A. & CHAN, A. T. 2020. Genetic Variant Associated With Survival of Patients With Stage II-III Colon Cancer. *Clinical Gastroenterology and Hepatology,* 18**,** 2717-2723.e3.

PERERA, D., CHACON, D., THOMS, J. A. I., POULOS, R. C., SHLIEN, A., BECK, D., CAMPBELL, P. J., PIMANDA, J. E. & WONG, J. W. H. 2014. OncoCis: annotation of cis-regulatory mutations in cancer. *Genome Biology,* 15**,** 485.

PÉREZ-HERRERO, E. & FERNÁNDEZ-MEDARDE, A. 2015. Advanced targeted therapies in cancer: Drug nanocarriers, the future of chemotherapy. *Eur J Pharm Biopharm,* 93**,** 52-79.

PETO, R., BOREHAM, J., LOPEZ, A. D., THUN, M. & HEATH, C. 1992. Mortality from tobacco in developed countries: indirect estimation from national vital statistics. *The Lancet,* 339**,** 1268-1278.

PHAN, L., JIN, Y., ZHANG, H., QIANG, W., SHEKHTMAN, E., SHAO, D., REVOE, D., VILLAMARIN, R., IVANCHENKO, E., KIMURA, M., WANG, Z., HAO, L., SHAROPOVA, N., BIHAN, M., STURCKE, A., LEE, M., POPOVA, N., WU, W., BASTIANI, C., WARD, M., HOLMES, J., LYOSHIN, V., KAUR, K., MOYER, E., FEOLO, M. & KATTMAN, B. 2020. *ALFA: Allele Frequency Aggregator* [Online]. National Center for Biotechnology Information, U.S. National Library of Medicine. Available: www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/ [Accessed 2020].

PHIPPS, A. I., BUCHANAN, D. D., MAKAR, K. W., WIN, A. K., BARON, J. A., LINDOR, N. M., POTTER, J. D. & NEWCOMB, P. A. 2013. KRAS-mutation status in relation to colorectal cancer survival: the joint impact of correlated tumour markers. *British Journal of Cancer,* 108**,** 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. & NEWCOMB, P. A. 2016. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *Carcinogenesis,* 37**,** 87-95.

PINO, M. S. & CHUNG, D. C. 2010. The chromosomal instability pathway in colon cancer. *Gastroenterology,* 138**,** 2059-72.

PLEGUEZUELOS-MANZANO, C., PUSCHHOF, J., ROSENDAHL HUBER, A., VAN HOECK, A., WOOD, H. M., NOMBURG, J., GURJAO, C., MANDERS, F., DALMASSO, G., STEGE, P. B., PAGANELLI, F. L., GEURTS, M. H., BEUMER, J., MIZUTANI, T., MIAO, Y., VAN DER LINDEN, R., VAN DER ELST, S., AMBROSE, J. C., ARUMUGAM, P., BAPLE, E. L., BLEDA, M., BOARDMAN-PRETTY, F., BOISSIERE, J. M., BOUSTRED, C. R., BRITTAIN, H., CAULFIELD, M. J., CHAN, G. C., CRAIG, C. E. H., DAUGHERTY, L. C., DE BURCA, A., DEVEREAU, A., ELGAR, G., FOULGER, R. E., FOWLER, T., FURIÓ-TARÍ, P., HACKETT, J. M., HALAI, D., HAMBLIN, A., HENDERSON, S., HOLMAN, J. E., HUBBARD, T. J. P., IBÁÑEZ, K., JACKSON, R., JONES, L. J., KASPERAVICIUTE, D., KAYIKCI, M., LAHNSTEIN, L., LAWSON, L., LEIGH, S. E. A., LEONG, I. U. S., LOPEZ, F. J., MALEADY-CROWE, F., MASON, J., MCDONAGH, E. M., MOUTSIANAS, L., MUELLER, M., MURUGAESU, N., NEED, A. C., ODHAMS, C. A., PATCH, C., PEREZ-GIL, D., POLYCHRONOPOULOS, D., PULLINGER, J., RAHIM, T., RENDON, A., RIESGO-FERREIRO, P., ROGERS, T., RYTEN, M., SAVAGE, K., SAWANT, K., SCOTT, R. H.,

SIDDIQ, A., SIEGHART, A., SMEDLEY, D., SMITH, K. R., SOSINSKY, A., SPOONER, W., STEVENS, H. E., STUCKEY, A., SULTANA, R., THOMAS, E. R. A., THOMPSON, S. R., TREGIDGO, C., TUCCI, A., WALSH, E., WATTERS, S. A., WELLAND, M. J., WILLIAMS, E., WITKOWSKA, K., WOOD, S. M., ZAROWIECKI, M., GARCIA, K. C., TOP, J., WILLEMS, R. J. L., GIANNAKIS, M., BONNET, R., QUIRKE, P., MEYERSON, M., CUPPEN, E., VAN BOXTEL, R., et al. 2020. Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. *Nature,* 580**,** 269-273.

POCIOT, F. 2017. Type 1 diabetes genome-wide association studies: not to be lost in translation. *Clin Transl Immunology,* 6**,** e162.

POMERANTZ, M. M. & FREEDMAN, M. L. 2011. The genetics of cancer risk. *Cancer J,* 17**,** 416-22.

POPAT, S., HUBNER, R. & HOULSTON, R. S. 2005. Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol,* 23**,** 609-18.

POTAPOVA, T. A., ZHU, J. & LI, R. 2013. Aneuploidy and chromosomal instability: a vicious cycle driving cellular evolution and cancer genome chaos. *Cancer Metastasis Rev,* 32**,** 377-89.

POULSEN, M. L. & BISGAARD, M. L. 2008. MUTYH Associated Polyposis (MAP). *Curr Genomics,* 9**,** 420-35.

POWELL, S. M., ZILZ, N., BEAZER-BARCLAY, Y., BRYAN, T. M., HAMILTON, S. R., THIBODEAU, S. N., VOGELSTEIN, B. & KINZLER, K. W. 1992. APC mutations occur early during colorectal tumorigenesis. *Nature,* 359**,** 235-7.

PRUIM, R. J., WELCH, R. P., SANNA, S., TESLOVICH, T. M., CHINES, P. S., GLIEDT, T. P., BOEHNKE, M., ABECASIS, G. R. & WILLER, C. J. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics,* 26**,** 2336-7.

PUBLIC HEALTH ENGLAND. 2020. *Young person and adult screening KPI data: annual (April 2018 to March 2019)* [Online]. Gov.uk. Available: https://www.gov.uk/government/publications/nhs-screening-programmes-kpi-reports-2018-to-2019 [Accessed 1st May 2021].

PUBLIC HEALTH WALES. 2020. *Bowel Screening Wales Annual Statisrtical Report 2018-2019* [Online]. Available: http://www.bowelscreening.wales.nhs.uk/sitesplus/documents/1121/Bowel%20Screening%20Wales%20Annual%20Statistical%20Report%202018-19%20v1.0.pdf [Accessed 1st May 2021].

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I., DALY, M. J. & SHAM, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81**,** 559-75.

QU, H. & FANG, X. 2013. A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. *Genomics Proteomics Bioinformatics,* 11**,** 135-41.

QU, S., YANG, X., LI, X., WANG, J., GAO, Y., SHANG, R., SUN, W., DOU, K. & LI, H. 2015. Circular RNA: A new star of noncoding RNAs. *Cancer Lett,* 365**,** 141-8.

QUYN, A. J., FRASER, C. G., STANNERS, G., CAREY, F. A., CARDEN, C., SHAUKAT, A. & STEELE, R. J. 2018. Uptake trends in the Scottish Bowel Screening Programme and the influences of age, sex, and deprivation. *J Med Screen,* 25**,** 24-31.

R CORE TEAM 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing

RAFIEMANESH, H., MOHAMMADIAN-HAFSHEJANI, A., GHONCHEH, M., SEPEHRI, Z., SHAMLOU, R., SALEHINIYA, H., TOWHIDI, F. & MAKHSOSI, B. R. 2016. Incidence and Mortality of Colorectal Cancer and Relationships with the Human Development Index across the World. *Asian Pac J Cancer Prev,* 17**,** 2465-73.

RANGANATHAN, P., PRAMESH, C. S. & BUYSE, M. 2016. Common pitfalls in statistical analysis: The perils of multiple testing. *Perspect Clin Res,* 7**,** 106-7.

RANIERI, G., PATRUNO, R., RUGGIERI, E., MONTEMURRO, S., VALERIO, P. & RIBATTI, D. 2006. Vascular endothelial growth factor (VEGF) as a target of bevacizumab in cancer: from the biology to the clinic. *Curr Med Chem,* 13**,** 1845-57.

RAWLA, P., SUNKARA, T. & BARSOUK, A. 2019. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Prz Gastroenterol,* 14**,** 89-103.

RAYCHAUDHURI, S. 2011. Mapping rare and common causal alleles for complex human diseases. *Cell,* 147**,** 57-69.

RAYMOND, E., CHANEY, S. G., TAAMMA, A. & CVITKOVIC, E. 1998. Oxaliplatin: a review of preclinical and clinical studies. *Ann Oncol,* 9**,** 1053-71.

RÉDA, M., RICHARD, C., BERTAUT, A., NIOGRET, J., COLLOT, T., FUMET, J. D., BLANC, J., TRUNTZER, C., DESMOULINS, I., LADOIRE, S., HENNEQUIN, A., FAVIER, L., BENGRINE, L., VINCENT, J., HERVIEU, A., DUSSERRE, J. G., LEPAGE, C., FOUCHER, P., BORG, C., ALBUISSON, J., ARNOULD, L., NAMBOT, S., FAIVRE, L., BOIDOT, R. & GHIRINGHELLI, F. 2020. Implementation and use of whole exome sequencing for metastatic solid cancer. *EBioMedicine,* 51**,** 102624.

REED, E., NUNEZ, S., KULP, D., QIAN, J., REILLY, M. P. & FOULKES, A. S. 2015. A guide to genome-wide association analysis and post-analytic interrogation. *Stat Med,* 34**,** 3769-92.

REICH, D. E. & LANDER, E. S. 2001. On the allelic spectrum of human disease. *Trends Genet,* 17**,** 502-10.

REIS, L. M., SOROKINA, E. A., THOMPSON, S., MUHEISEN, S., VELINOV, M., ZAMORA, C., AYLSWORTH, A. S. & SEMINA, E. V. 2019. De Novo Missense Variants in WDR37 Cause a Severe Multisystemic Syndrome. *Am J Hum Genet,* 105**,** 425-433.

RENTSCH, M., SCHIERGENS, T., KHANDOGA, A. & WERNER, J. 2016. Surgery for Colorectal Cancer - Trends, Developments, and Future Perspectives. *Visc Med,* 32**,** 184-91.

RHOADES, R., JACKSON, F. & TENG, S. 2019. Discovery of rare variants implicated in schizophrenia using next-generation sequencing. *J Transl Genet Genom,* 3**,** 1-20.

RICE, T. K., SCHORK, N. J. & RAO, D. C. 2008. Methods for handling multiple testing. *Adv Genet,* 60**,** 293-308.

RICHARD, G. F., KERREST, A. & DUJON, B. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev,* 72**,** 686-727.

RICHMAN, S. D., SEYMOUR, M. T., CHAMBERS, P., ELLIOTT, F., DALY, C. L., MEADE, A. M., TAYLOR, G., BARRETT, J. H. & 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. *J Clin Oncol,* 27**,** 5931-7.

RIERA ROMO, M., PEREZ-MARTINEZ, D. & CASTILLO FERRER, C. 2016. Innate immunity in vertebrates: an overview. *Immunology,* 148**,** 125-39.

RIGO, A., GOTTARDI, M., ZAMÒ, A., MAURI, P., BONIFACIO, M., KRAMPERA, M., DAMIANI, E., PIZZOLO, G. & VINANTE, F. 2010. Macrophages may promote cancer growth via a GM-CSF/HB-EGF paracrine loop that is enhanced by CXCL12. *Mol Cancer,* 9**,** 273.

RIIHIMÄKI, M., HEMMINKI, A., SUNDQUIST, J. & HEMMINKI, K. 2016. Patterns of metastasis in colon and rectal cancer. *Sci Rep,* 6**,** 29765.

RISCH, N. & MERIKANGAS, K. 1996. The future of genetic studies of complex human diseases. *Science,* 273**,** 1516-7.

RISCH, N. J. 2000. Searching for genetic determinants in the new millennium. *Nature,* 405**,** 847-856.

ROBERTS, N. A. 2015. The Immune System. *The Yale Journal of Biology and Medicine,* 88**,** 99-99.

ROMAGNOLO, B., BERREBI, D., SAADI-KEDDOUCCI, S., PORTEU, A., PICHARD, A. L., PEUCHMAUR, M., VANDEWALLE, A., KAHN, A. & PERRET, C. 1999. Intestinal dysplasia and adenoma in transgenic mice after overexpression of an activated beta-catenin. *Cancer Res,* 59**,** 3875-9.

ROSE, M. G., FARRELL, M. P. & SCHMITZ, J. C. 2002. Thymidylate synthase: a critical target for cancer chemotherapy. *Clin Colorectal Cancer,* 1**,** 220-9.

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. & 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**,** 1031-9.

ROTHENBERG, M. L. 2000. Efficacy of oxaliplatin in the treatment of colorectal cancer. *Oncology (Williston Park),* 14**,** 9-14.

RSTUDIO TEAM 2019. RStudio: Integrated Development for R.

RUBARTELLI, A. & LOTZE, M. T. 2007. Inside, outside, upside down: damage-associated molecular-pattern molecules (DAMPs) and redox. *Trends Immunol,* 28**,** 429-36.

RUSSO, A., BAZAN, V., IACOPETTA, B., KERR, D., SOUSSI, T. & 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**,** 7518-28.

SAKHARKAR, M. K., CHOW, V. T. & KANGUEANE, P. 2004. Distributions of exons and introns in the human genome. *In Silico Biol,* 4**,** 387-93.

SALAMA, P. & PLATELL, C. 2008. Host response to colorectal cancer. *ANZ J Surg,* 78**,** 745-53.

SALCEDO, R., WORSCHECH, A., CARDONE, M., JONES, Y., GYULAI, Z., DAI, R. M., WANG, E., MA, W., HAINES, D., O'HUIGIN, C., MARINCOLA, F. M. & TRINCHIERI, G. 2010. MyD88-mediated signaling prevents development of adenocarcinomas of the colon: role of interleukin 18. *J Exp Med,* 207**,** 1625-36.

SALZBERG, S. L. 2018. Open questions: How many genes do we have? *BMC Biol,* 16**,** 94.

SAMOWITZ, W. S., SWEENEY, C., HERRICK, J., ALBERTSEN, H., LEVIN, T. R., MURTAUGH, M. A., WOLFF, R. K. & SLATTERY, M. L. 2005. Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers. *Cancer Res,* 65**,** 6063-9.

SANSREGRET, L., VANHAESEBROECK, B. & SWANTON, C. 2018. Determinants and clinical implications of chromosomal instability in cancer. *Nat Rev Clin Oncol,* 15**,** 139-150.

SANTI, D. V. & HARDY, L. W. 1987. Catalytic mechanism and inhibition of tRNA (uracil-5-)methyltransferase: evidence for covalent catalysis. *Biochemistry,* 26**,** 8599-606.

SANTOS, M. F., MANNAM, V. K., CRAFT, B. S., PUNEKY, L. V., SHEEHAN, N. T., LEWIS, R. E. & CRUSE, J. M. 2014. Comparative analysis of innate immune system function in metastatic breast, colorectal, and prostate cancer patients with circulating tumor cells. *Exp Mol Pathol,* 96**,** 367-74.

SANZ-GARCIA, E., ARGILES, G., ELEZ, E. & TABERNERO, J. 2017. BRAF mutant colorectal cancer: prognosis, treatment, and new perspectives. *Ann Oncol,* 28**,** 2648-2657.

SCARTOZZI, M., GIAMPIERI, R., APRILE, G., IACONO, D., SANTINI, D., DELL'AQUILA, E., SILVESTRIS, N., GNONI, A., BONOTTO, M., PUZZONI, M., DEMURTAS, L. & CASCINU, S. 2015. The distinctive molecular, pathological and clinical characteristics of BRAF-mutant colorectal tumors. *Expert Rev Mol Diagn,* 15**,** 979-87.

SCHATOFF, E. M., LEACH, B. I. & DOW, L. E. 2017. Wnt Signaling and Colorectal Cancer. *Curr Colorectal Cancer Rep,* 13**,** 101-110.

SCHEITHAUER, W., ROSEN, H., KORNEK, G. V., SEBESTA, C. & DEPISCH, D. 1993. Randomised comparison of combination chemotherapy plus supportive care with supportive care alone in patients with metastatic colorectal cancer. *Bmj,* 306**,** 752-5.

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. & FALCONE, A. 2015. Role of NRAS mutations as prognostic and predictive markers in metastatic colorectal cancer. *Int J Cancer,* 136**,** 83-90.

SCHLUSSEL, A. T., GAGLIANO JR, R. A., SETO-DONLON, S., EGGERDING, F., DONLON, T., BERENBERG, J. & LYNCH, H. T. 2014. The evolution of colorectal cancer genetics—Part 1: from discovery to practice. *Journal of Gastrointestinal Oncology,* 5**,** 326-335.

SCHMUCK, R., GERKEN, M., TEEGEN, E. M., KREBS, I., KLINKHAMMER-SCHALKE, M., AIGNER, F., PRATSCHKE, J., RAU, B. & BENZ, S. 2020. Gender comparison of clinical, histopathological, therapeutic and outcome

factors in 185,967 colon cancer patients. *Langenbecks Arch Surg,* 405**,** 71-80.
SCHNEIKERT, J. & BEHRENS, J. 2007. The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut,* 56**,** 417-25.
SCHOENFELDER, S. & FRASER, P. 2019. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics,* 20**,** 437-455.
SCHRODER, K. & TSCHOPP, J. 2010. The inflammasomes. *Cell,* 140**,** 821-32.
SCHRÖDER, N. W. J. & SCHUMANN, R. R. 2005. Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *The Lancet Infectious Diseases,* 5**,** 156-164.
SCHUURHUIZEN, C. S., VERHEUL, H. M., BRAAMSE, A. M., BUFFART, L. M., BLOEMENDAL, H. J., DEKKER, J. & KONINGS, I. R. 2018. The predictive value of cumulative toxicity for quality of life in patients with metastatic colorectal cancer during first-line palliative chemotherapy. *Cancer Manag Res,* 10**,** 3015-3021.
SCHWARZE, K., BUCHANAN, J., TAYLOR, J. C. & WORDSWORTH, S. 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genet Med,* 20**,** 1122-1130.
SCHWINGSHACKL, L., SCHWEDHELM, C., HOFFMANN, G., KNÜPPEL, S., LAURE PRETERRE, A., IQBAL, K., BECHTHOLD, A., DE HENAUW, S., MICHELS, N., DEVLEESSCHAUWER, B., BOEING, H. & SCHLESINGER, S. 2018. Food groups and risk of colorectal cancer. *Int J Cancer,* 142**,** 1748-1758.
SEABY, E. G., PENGELLY, R. J. & ENNIS, S. 2016. Exome sequencing explained: a practical guide to its clinical application. *Briefings in Functional Genomics,* 15**,** 374-384.
SEGELOV, E., WARING, P., DESAI, J., WILSON, K., GEBSKI, V., THAVANESWARAN, S., ELEZ, E., UNDERHILL, C., PAVLAKIS, N., CHANTRILL, L., NOTT, L., JEFFORD, M., KHASRAW, M., DAY, F., WASAN, H., CIARDIELLO, F., KARAPETIS, C., JOUBERT, W., VAN HAZEL, G., HAYDON, A., PRICE, T., TEJPAR, S., TEBBUTT, N. & SHAPIRO, J. 2016. ICECREAM: randomised phase II study of cetuximab alone or in combination with irinotecan in patients with metastatic colorectal cancer with either KRAS, NRAS, BRAF and PI3KCA wild type, or G13D mutated tumours. *BMC cancer,* 16**,** 339-339.
SEIFERT, R. & WENZEL-SEIFERT, K. 2003. The human formyl peptide receptor as model system for constitutively active G-protein-coupled receptors. *Life Sci,* 73**,** 2263-80.
SEY, N. Y. A., HU, B., MAH, W., FAUNI, H., MCAFEE, J. C., RAJARAJAN, P., BRENNAND, K. J., AKBARIAN, S. & WON, H. 2020. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci,* 23**,** 583-593.
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. & 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**,** 749-59.

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. & 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**,** 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. & LANGLEY, R. E. 2011. Chemotherapy options in elderly and frail patients with metastatic colorectal cancer (MRC FOCUS2): an open-label, randomised factorial trial. *The Lancet,* 377**,** 1749-1759.

SHAM, P. C. & PURCELL, S. M. 2014. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics,* 15**,** 335-346.

SHARMA, N., SAIFO, M., TAMASKAR, I. R., BHUVANESWARI, R., MASHTARE, T. & FAKIH, M. 2010. KRAS status and clinical outcome in metastatic colorectal cancer patients treated with first-line FOLFOX chemotherapy. *Journal of gastrointestinal oncology,* 1**,** 90-96.

SHASTRY, B. S. 2009. SNPs: Impact on Gene Function and Phenotype. *In:* KOMAR, A. A. (ed.) *Single Nucleotide Polymorphisms: Methods and Protocols.* Totowa, NJ: Humana Press.

SHAUKAT, A., DOSTAL, A., MENK, J. & CHURCH, T. R. 2017. BMI Is a Risk Factor for Colorectal Cancer Mortality. *Digestive Diseases and Sciences,* 62**,** 2511-2517.

SHAW, E., FARRIS, M. S., STONE, C. R., DERKSEN, J. W. G., JOHNSON, R., HILSDEN, R. J., FRIEDENREICH, C. M. & BRENNER, D. R. 2018. Effects of physical activity on colorectal cancer risk among family history and body mass index subgroups: a systematic review and meta-analysis. *BMC Cancer,* 18**,** 71.

SHERRY, S. T., WARD, M. H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E. M. & SIROTKIN, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res,* 29**,** 308-11.

SHORTMAN, K. & LIU, Y. J. 2002. Mouse and human dendritic cell subtypes. *Nat Rev Immunol,* 2**,** 151-61.

SIEGEL, R. L., JEMAL, A., THUN, M. J., HAO, Y. & WARD, E. M. 2008. Trends in the incidence of colorectal cancer in relation to county-level poverty among blacks and whites. *J Natl Med Assoc,* 100**,** 1441-4.

SIGISMUND, S., AVANZATO, D. & LANZETTI, L. 2018. Emerging functions of the EGFR in cancer. *Mol Oncol,* 12**,** 3-20.

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. & 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**,** 1843-52.

SIMON, R. & NORTON, L. 2006. The Norton–Simon hypothesis: designing more effective and less toxic chemotherapeutic regimens. *Nature Clinical Practice Oncology,* 3**,** 406-407.

SINICROPE, F. A. & SARGENT, D. J. 2012. Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications. *Clin Cancer Res,* 18**,** 1506-12.

SISKA, P. J. & RATHMELL, J. C. 2015. T cell metabolic fitness in antitumor immunity. *Trends Immunol,* 36**,** 257-64.

SKOPELITOU, D., MIAO, B., SRIVASTAVA, A., KUMAR, A., KUSWICK, M., DYMERSKA, D., PARAMASIVAM, N., SCHLESNER, M., LUBINSKI, J., HEMMINKI, K., FÖRSTI, A. & BANDAPALLI, O. R. 2021. Whole Exome Sequencing Identifies APCDD1 and HDAC5 Genes as Potentially Cancer Predisposing in Familial Colorectal Cancer. *Int J Mol Sci,* 22.

SLATKIN, M. 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics,* 9**,** 477-485.

SLATTERY, M. L., HERRICK, J. S., MULLANY, L. E., SAMOWITZ, W. S., SEVENS, J. R., SAKODA, L. & WOLFF, R. K. 2017. The co-regulatory networks of tumor suppressor genes, oncogenes, and miRNAs in colorectal cancer. *Genes Chromosomes Cancer,* 56**,** 769-787.

SLOAN, C. A., CHAN, E. T., DAVIDSON, J. M., MALLADI, V. S., STRATTAN, J. S., HITZ, B. C., GABDANK, I., NARAYANAN, A. K., HO, M., LEE, B. T., ROWE, L. D., DRESZER, T. R., ROE, G., PODDUTURI, N. R., TANAKA, F., HONG, E. L. & CHERRY, J. M. 2016. ENCODE data at the ENCODE portal. *Nucleic Acids Res,* 44**,** D726-32.

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. & CHEADLE, J. P. 2013a. 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**,** 4104-13.

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. & CHEADLE, J. P. 2013b. 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**,** 4104-13.

SMITH, C. G., FISHER, D., HARRIS, R., MAUGHAN, T. S., PHIPPS, A. I., RICHMAN, S., SEYMOUR, M., TOMLINSON, I., ROSMARIN, D., KERR, D., CHAN, A. T., PETERS, U., NEWCOMB, P. A., IDZIASZCZYK, S., WEST, H., MEADE, A., KAPLAN, R. & 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,* 21**,** 3453-61.

SMITH, J. C., BROOKS, L., HOFF, P. M., MCWALTER, G., DEARDEN, S., MORGAN, S. R., WILSON, D., ROBERTSON, J. D. & JURGENSMEIER, J. M. 2013c. KRAS mutations are associated with inferior clinical outcome in patients with metastatic colorectal cancer, but are not predictive for benefit with cediranib. *Eur J Cancer,* 49**,** 2424-32.

SNYDER, C. & HAMPEL, H. 2019. Hereditary Colorectal Cancer Syndromes. *Semin Oncol Nurs,* 35**,** 58-78.

SOBRERO, A., LONARDI, S., ROSATI, G., DI BARTOLOMEO, M., RONZONI, M., PELLA, N., SCARTOZZI, M., BANZI, M., ZAMPINO, M. G., PASINI, F., MARCHETTI, P., CANTORE, M., ZANIBONI, A., RIMASSA, L., CIUFFREDA, L., FERRARI, D., ZAGONEL, V., MAIELLO, E., BARNI, S., RULLI, E. & LABIANCA, R. 2018. FOLFOX or CAPOX in Stage II to III Colon Cancer: Efficacy Results of the Italian Three or Six Colon Adjuvant Trial. *J Clin Oncol,* 36**,** 1478-1485.

SOUGLAKOS, J., PHILIPS, J., WANG, R., MARWAH, S., SILVER, M., TZARDI, M., SILVER, J., OGINO, S., HOOSHMAND, S., KWAK, E., FREED, E., MEYERHARDT, J. A., SARIDAKI, Z., GEORGOULIAS, V., FINKELSTEIN, D., FUCHS, C. S., KULKE, M. H. & SHIVDASANI, R. A. 2009. Prognostic and predictive value of common mutations for treatment response and survival in patients with metastatic colorectal cancer. *Br J Cancer,* 101**,** 465-72.

STEELE, S. R., PARK, G. E., JOHNSON, E. K., MARTIN, M. J., STOJADINOVIC, A., MAYKEL, J. A. & CAUSEY, M. W. 2014. The impact of age on colorectal cancer incidence, treatment, and outcomes in an equal-access health care system. *Dis Colon Rectum,* 57**,** 303-10.

STEINBERG, S. M., BARKIN, J. S., KAPLAN, R. S. & STABLEIN, D. M. 1986. Prognostic indicators of colon tumors. The Gastrointestinal Tumor Study Group experience. *Cancer,* 57**,** 1866-70.

STINTZING, S., MODEST, D. P., ROSSIUS, L., LERCH, M. M., VON WEIKERSTHAL, L. F., DECKER, T., KIANI, A., VEHLING-KAISER, U., AL-BATRAN, S. E., HEINTGES, T., LERCHENMÜLLER, C., KAHL, C., SEIPELT, G., KULLMANN, F., STAUCH, M., SCHEITHAUER, W., HELD, S., GIESSEN-JUNG, C., MOEHLER, M., JAGENBURG, A., KIRCHNER, T., JUNG, A. & HEINEMANN, V. 2016. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab for metastatic colorectal cancer (FIRE-3): a post-hoc analysis of tumour dynamics in the final RAS wild-type subgroup of this randomised open-label phase 3 trial. *Lancet Oncol,* 17**,** 1426-1434.

STOREY, J., BASS, A., DABNEY, A. & ROBINSON, D. 2020. *qvalue: Q-value estimation for false discovery rate control*. R package version 2.20.0.

SU, L. K., VOGELSTEIN, B. & KINZLER, K. W. 1993. Association of the APC tumor suppressor protein with catenins. *Science,* 262**,** 1734-7.

SUMMERS, M. G. 2019. *Searching for genetic variants associated with survival in patients with metastatic colorectal cancer.* Doctor of Philosophy (PhD), Cardiff University.

SUNG, H., FERLAY, J., SIEGEL, R. L., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*.

SUWINSKI, P., ONG, C., LING, M. H. T., POH, Y. M., KHAN, A. M. & ONG, H. S. 2019. Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics. *Frontiers in Genetics,* 10.

SUZUKI, C., BLOMQVIST, L., SUNDIN, A., JACOBSSON, H., BYSTRÖM, P., BERGLUND, Å., NYGREN, P. & GLIMELIUS, B. 2012. The initial change in tumor size predicts response and survival in patients with metastatic colorectal cancer treated with combination chemotherapy. *Ann Oncol,* 23**,** 948-54.

SZYMCZAK, S., BIERNACKA, J. M., CORDELL, H. J., GONZÁLEZ-RECIO, O., KÖNIG, I. R., ZHANG, H. & SUN, Y. V. 2009. Machine learning in genome-wide association studies. *Genet Epidemiol,* 33 Suppl 1**,** S51-7.

T LAM-BOER, J., MOL, L., VERHOEF, C., DE HAAN, A. F., YILMAZ, M., PUNT, C. J., DE WILT, J. H. & 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.

TAKEUCHI, F., YANAI, K., MORII, T., ISHINAGA, Y., TANIGUCHI-YANAI, K., NAGANO, S. & KATO, N. 2005. Linkage disequilibrium grouping of single nucleotide polymorphisms (SNPs) reflecting haplotype phylogeny for efficient selection of tag SNPs. *Genetics,* 170**,** 291-304.

TAKEUCHI, O. & AKIRA, S. 2010. Pattern recognition receptors and inflammation. *Cell,* 140**,** 805-20.

TANIKAWA, C., KAMATANI, Y., TOYOSHIMA, O., SAKAMOTO, H., ITO, H., TAKAHASHI, A., MOMOZAWA, Y., HIRATA, M., FUSE, N., TAKAI-IGARASHI, T., SHIMIZU, A., SASAKI, M., YAMAJI, T., SAWADA, N., IWASAKI, M., TSUGANE, S., NAITO, M., HISHIDA, A., WAKAI, K., FURUSYO, N., MURAKAMI, Y., NAKAMURA, Y., IMOTO, I., INAZAWA, J., OZE, I., SATO, N., TANIOKA, F., SUGIMURA, H., HIROSE, H., YOSHIDA, T., MATSUO, K., KUBO, M. & MATSUDA, K. 2018. Genome-wide association study identifies gastric cancer susceptibility loci at 12q24.11-12 and 20q11.21. *Cancer Sci,* 109**,** 4015-4024.

TAPIAL, S., OLMEDILLAS-LÓPEZ, S., RUEDA, D., ARRIBA, M., GARCÍA, J. L., VIVAS, A., PÉREZ, J., PENA-COUSO, L., OLIVERA, R., RODRÍGUEZ, Y., GARCÍA-ARRANZ, M., GARCÍA-OLMO, D., GONZÁLEZ-SARMIENTO, R., URIOSTE, M., GOEL, A. & PEREA, J. 2019. Cimp-Positive Status is More Representative in Multiple Colorectal Cancers than in Unique Primary Colorectal Cancers. *Scientific Reports,* 9**,** 10516.

TAY, R. E., RICHARDSON, E. K. & TOH, H. C. 2021. Revisiting the role of CD4+ T cells in cancer immunotherapy—new insights into old paradigms. *Cancer Gene Therapy,* 28**,** 5-17.

TENNESSEN, J. A., BIGHAM, A. W., O'CONNOR, T. D., FU, W., KENNY, E. E., GRAVEL, S., MCGEE, S., DO, R., LIU, X., JUN, G., KANG, H. M., JORDAN, D., LEAL, S. M., GABRIEL, S., RIEDER, M. J., ABECASIS, G., ALTSHULER, D., NICKERSON, D. A., BOERWINKLE, E., SUNYAEV, S., BUSTAMANTE, C. D., BAMSHAD, M. J. & AKEY, J. M. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science,* 337**,** 64-9.

TESNIERE, A., SCHLEMMER, F., BOIGE, V., KEPP, O., MARTINS, I., GHIRINGHELLI, F., AYMERIC, L., MICHAUD, M., APETOH, L., BARAULT, L., MENDIBOURE, J., PIGNON, J. P., JOOSTE, V., VAN ENDERT, P., DUCREUX, M., ZITVOGEL, L., PIARD, F. & KROEMER, G. 2010. Immunogenic death of colon cancer cells treated with oxaliplatin. *Oncogene,* 29**,** 482-91.

THE ENCODE PROJECT CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

THE GTEX CONSORTIUM 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet,* 45**,** 580-5.

THERASSE, P., ARBUCK, S. G., EISENHAUER, E. A., WANDERS, J., KAPLAN, R. S., RUBINSTEIN, L., VERWEIJ, J., VAN GLABBEKE, M., VAN OOSTEROM, A. T., CHRISTIAN, M. C. & GWYTHER, S. G. 2000. 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,* 92**,** 205-16.

THERNEAU, T. M. 2020. A Package for Survival Analysis in R.

THOMPSON, S. L., BAKHOUM, S. F. & COMPTON, D. A. 2010. Mechanisms of chromosomal instability. *Curr Biol,* 20**,** R285-95.

TIAN, J., CAI, Y., LI, Y., LU, Z., HUANG, J., DENG, Y., YANG, N., WANG, X., YING, P., ZHANG, S., ZHU, Y., ZHANG, H., ZHONG, R., CHANG, J. & MIAO, X. 2021. CancerImmunityQTL: a database to systematically evaluate the impact of genetic variants on immune infiltration in human cancer. *Nucleic Acids Research,* 49**,** D1065-D1073.

TIAN, J., LOU, J., CAI, Y., RAO, M., LU, Z., ZHU, Y., ZOU, D., PENG, X., WANG, H., ZHANG, M., NIU, S., LI, Y., ZHONG, R., CHANG, J. & MIAO, X. 2020. Risk SNP-Mediated Enhancer-Promoter Interaction Drives Colorectal Cancer through Both FADS2 and AP002754.2. *Cancer Res,* 80**,** 1804-1818.

TODD, R. C. & LIPPARD, S. J. 2009. Inhibition of transcription by platinum antitumor compounds. *Metallomics,* 1**,** 280-91.

TOL, J., KOOPMAN, M., CATS, A., RODENBURG, C. J., CREEMERS, G. J., SCHRAMA, J. G., ERDKAMP, F. L., VOS, A. H., VAN GROENINGEN, C. J., SINNIGE, H. A., RICHEL, D. J., VOEST, E. E., DIJKSTRA, J. R., VINK-BÖRGER, M. E., ANTONINI, N. F., MOL, L., VAN KRIEKEN, J. H., DALESIO, O. & PUNT, C. J. 2009. Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. *N Engl J Med,* 360**,** 563-72.

TOL, J., KOOPMAN, M., RODENBURG, C. J., CATS, A., CREEMERS, G. J., SCHRAMA, J. G., ERDKAMP, F. L., VOS, A. H., MOL, L., ANTONINI, N. F. & PUNT, C. J. 2008. A randomised phase III study on capecitabine, oxaliplatin and bevacizumab with or without cetuximab in first-line advanced colorectal cancer, the CAIRO2 study of the Dutch Colorectal Cancer Group (DCCG). An interim analysis of toxicity. *Ann Oncol,* 19**,** 734-8.

TORRES STONE, R. A., WARING, M. E., CUTRONA, S. L., KIEFE, C. I., ALLISON, J. & DOUBENI, C. A. 2017. The association of dietary quality with colorectal cancer among normal weight, overweight and obese men and women: a prospective longitudinal study in the USA. *BMJ Open,* 7**,** e015619.

TORRY, D. S. & COOPER, G. M. 1991. Proto-oncogenes in development and cancer. *Am J Reprod Immunol,* 25**,** 129-32.

TOSOLINI, M., KIRILOVSKY, A., MLECNIK, B., FREDRIKSEN, T., MAUGER, S., BINDEA, G., BERGER, A., BRUNEVAL, P., FRIDMAN, W. H., PAGES, F. & GALON, J. 2011. Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, th2, treg, th17) in patients with colorectal cancer. *Cancer Res,* 71**,** 1263-71.

TRAN, B., KOPETZ, S., TIE, J., GIBBS, P., JIANG, Z. Q., LIEU, C. H., AGARWAL, A., MARU, D. M., SIEBER, O. & DESAI, J. 2011. Impact of BRAF mutation and microsatellite instability on the pattern of metastatic spread and prognosis in metastatic colorectal cancer. *Cancer,* 117**,** 4623-32.

TSUKUMO, S. I. & YASUTOMO, K. 2018. Regulation of CD8(+) T Cells and Antitumor Immunity by Notch Signaling. *Front Immunol,* 9**,** 101.

TULAH, A. S., HOLLOWAY, J. W. & SAYERS, I. 2013. Defining the contribution of SNPs identified in asthma GWAS to clinical variables in asthmatic children. *BMC Med Genet,* 14**,** 100.

TURAJLIC, S., SOTTORIVA, A., GRAHAM, T. & SWANTON, C. 2019. Resolving genetic heterogeneity in cancer. *Nature Reviews Genetics,* 20**,** 404-416.

TURNBULL, C., SCOTT, R. H., THOMAS, E., JONES, L., MURUGAESU, N., PRETTY, F. B., HALAI, D., BAPLE, E., CRAIG, C., HAMBLIN, A., HENDERSON, S., PATCH, C., O'NEILL, A., DEVEREAU, A., SMITH, K., MARTIN, A. R., SOSINSKY, A., MCDONAGH, E. M., SULTANA, R., MUELLER, M., SMEDLEY, D., TOMS, A., DINH, L., FOWLER, T., BALE, M., HUBBARD, T., RENDON, A., HILL, S. & CAULFIELD, M. J. 2018. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *Bmj,* 361**,** k1687.

TURNER, S. D. 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv***,** 005165.

TWELVES, C., BOYER, M., FINDLAY, M., CASSIDY, J., WEITZEL, C., BARKER, C., OSTERWALDER, B., JAMIESON, C. & HIEKE, K. 2001. Capecitabine (Xeloda) improves medical resource use compared with 5-fluorouracil plus leucovorin in a phase III trial conducted in patients with advanced colorectal carcinoma. *Eur J Cancer.* England.

UDDEN, S. M. N., PENG, L., GAN, J. L., SHELTON, J. M., MALTER, J. S., HOOPER, L. V. & ZAKI, M. H. 2017. NOD2 Suppresses Colorectal Tumorigenesis via Downregulation of the TLR Pathways. *Cell Rep,* 19**,** 2756-2770.

UFFELMANN, E., HUANG, Q. Q., MUNUNG, N. S., DE VRIES, J., OKADA, Y., MARTIN, A. R., MARTIN, H. C., LAPPALAINEN, T. & POSTHUMA, D. 2021. Genome-wide association studies. *Nature Reviews Methods Primers,* 1**,** 59.

UHLEN, M., ZHANG, C., LEE, S., SJÖSTEDT, E., FAGERBERG, L., BIDKHORI, G., BENFEITAS, R., ARIF, M., LIU, Z., EDFORS, F., SANLI, K., VON FEILITZEN, K., OKSVOLD, P., LUNDBERG, E., HOBER, S., NILSSON, P., MATTSSON, J., SCHWENK, J. M., BRUNNSTRÖM, H., GLIMELIUS, B., SJÖBLOM, T., EDQVIST, P. H., DJUREINOVIC, D., MICKE, P., LINDSKOG, C., MARDINOGLU, A. & PONTEN, F. 2017. A pathology atlas of the human cancer transcriptome. *Science,* 357.

VACCHELLI, E., ENOT, D. P., PIETROCOLA, F., ZITVOGEL, L. & KROEMER, G. 2016. Impact of Pattern Recognition Receptors on the Prognosis of Breast Cancer Patients Undergoing Adjuvant Chemotherapy. *Cancer Res,* 76**,** 3122-6.

VACCHELLI, E., MA, Y., BARACCO, E. E., SISTIGU, A., ENOT, D. P., PIETROCOLA, F., YANG, H., ADJEMIAN, S., CHABA, K., SEMERARO, M., SIGNORE, M., DE NINNO, A., LUCARINI, V., PESCHIAROLI, F., BUSINARO, L., GERARDINO, A., MANIC, G., ULAS, T., GUNTHER, P., SCHULTZE, J. L., KEPP, O., STOLL, G., LEFEBVRE, C., MULOT, C., CASTOLDI, F., RUSAKIEWICZ, S., LADOIRE, S., APETOH, L., BRAVO-SAN PEDRO, J. M., LUCATTELLI, M., DELARASSE, C., BOIGE, V., DUCREUX, M., DELALOGE, S., BORG, C., ANDRE, F., SCHIAVONI, G., VITALE, I., LAURENT-PUIG, P., MATTEI, F., ZITVOGEL, L. & KROEMER,

G. 2015. Chemotherapy-induced antitumor immunity requires formyl peptide receptor 1. *Science,* 350**,** 972-8.

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., KÖHNE, C. H., LABIANCA, R., LAURENT-PUIG, P., MA, B., MAUGHAN, T., MURO, K., NORMANNO, N., ÖSTERLUND, P., OYEN, W. J. G., PAPAMICHAEL, D., PENTHEROUDAKIS, G., PFEIFFER, P., PRICE, T. J., PUNT, C., RICKE, J., ROTH, A., SALAZAR, R., SCHEITHAUER, W., SCHMOLL, H. J., TABERNERO, J., TAÏEB, J., TEJPAR, S., WASAN, H., YOSHINO, T., ZAANAN, A. & ARNOLD, D. 2016. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Annals of Oncology,* 27**,** 1386-1422.

VAN CUTSEM, E., CERVANTES, A., NORDLINGER, B. & 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., HOFF, P. M., HARPER, P., BUKOWSKI, R. M., CUNNINGHAM, D., DUFOUR, P., GRAEVEN, U., LOKICH, J., MADAJEWICZ, S., MAROUN, J. A., MARSHALL, J. L., MITCHELL, E. P., PEREZ-MANGA, G., ROUGIER, P., SCHMIEGEL, W., SCHOELMERICH, J., SOBRERO, A. & SCHILSKY, R. L. 2004. Oral capecitabine vs intravenous 5-fluorouracil and leucovorin: integrated efficacy data and novel analyses from two large, randomised, phase III trials. *Br J Cancer,* 90**,** 1190-7.

VAN CUTSEM, E., KÖHNE, C. H., HITRE, E., ZALUSKI, J., CHANG CHIEN, C. R., MAKHSON, A., D'HAENS, G., PINTÉR, T., LIM, R., BODOKY, G., ROH, J. K., FOLPRECHT, G., RUFF, P., STROH, C., TEJPAR, S., SCHLICHTING, M., NIPPGEN, J. & ROUGIER, P. 2009. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med,* 360**,** 1408-17.

VAN CUTSEM, E., KÖHNE, C. H., LÁNG, I., FOLPRECHT, G., NOWACKI, M. P., CASCINU, S., SHCHEPOTIN, I., MAUREL, J., CUNNINGHAM, D., TEJPAR, S., SCHLICHTING, M., ZUBEL, A., CELIK, I., ROUGIER, P. & 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**,** 2011-9.

VAN CUTSEM, E., MAYER, R. J., LAURENT, S., WINKLER, R., GRÁVALOS, C., BENAVIDES, M., LONGO-MUNOZ, F., PORTALES, F., CIARDIELLO, F., SIENA, S., YAMAGUCHI, K., MURO, K., DENDA, T., TSUJI, Y., MAKRIS, L., LOEHRER, P., LENZ, H. J. & OHTSU, A. 2018. The subgroups of the phase III RECOURSE trial of trifluridine/tipiracil (TAS-102) versus placebo with best supportive care in patients with metastatic colorectal cancer. *Eur J Cancer,* 90**,** 63-72.

VAN DER VELDEN, D. L., OPDAM, F. L. & OPDAM, F. L. 2017. TAS-102 and the quest for predictive biomarkers. *ESMO Open,* 2**,** e000263.

VAN EEGHEN, E. E., BAKKER, S. D., VAN BOCHOVE, A. & LOFFELD, R. J. 2015. Impact of age and comorbidity on survival in colorectal cancer. *J Gastrointest Oncol,* 6**,** 605-12.

VANLIERE, J. M. & ROSENBERG, N. A. 2008. Mathematical properties of the r2 measure of linkage disequilibrium. *Theor Popul Biol,* 74**,** 130-7.

VARLEY, K. E., GERTZ, J., BOWLING, K. M., PARKER, S. L., REDDY, T. E., PAULI-BEHN, F., CROSS, M. K., WILLIAMS, B. A., STAMATOYANNOPOULOS, J. A., CRAWFORD, G. E., ABSHER, D. M., WOLD, B. J. & MYERS, R. M. 2013. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res,* 23**,** 555-67.

VERDAGUER, H., SAURÍ, T. & MACARULLA, T. 2017. Predictive and prognostic biomarkers in personalized gastrointestinal cancer treatment. *J Gastrointest Oncol,* 8**,** 405-417.

VICENTE-DUEÑAS, C., ROMERO-CAMARERO, I., COBALEDA, C. & SÁNCHEZ-GARCÍA, I. 2013. Function of oncogenes in cancer development: a changing paradigm. *Embo j,* 32**,** 1502-13.

VIRAG, P., FISCHER-FODOR, E., PERDE-SCHREPLER, M., BRIE, I., TATOMIR, C., BALACESCU, L., BERINDAN-NEAGOE, I., VICTOR, B. & BALACESCU, O. 2013. Oxaliplatin induces different cellular and molecular chemoresistance patterns in colorectal cancer cell lines of identical origins. *BMC Genomics,* 14**,** 480.

VODENKOVA, S., BUCHLER, T., CERVENA, K., VESKRNOVA, V., VODICKA, P. & VYMETALKOVA, V. 2020. 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. *Pharmacol Ther,* 206**,** 107447.

VODICKA, P., VODENKOVA, S., BUCHLER, T. & VODICKOVA, L. 2019. DNA repair capacity and response to treatment of colon cancer. *Pharmacogenomics,* 20**,** 1225-1233.

VON WAGNER, C., GOOD, A., WRIGHT, D., RACHET, B., OBICHERE, A., BLOOM, S. & WARDLE, J. 2009. Inequalities in colorectal cancer screening participation in the first round of the national screening programme in England. *British Journal of Cancer,* 101**,** S60-S63.

WALDNER, M., SCHIMANSKI, C. C. & NEURATH, M. F. 2006. Colon cancer and the immune system: the role of tumor invading T cells. *World J Gastroenterol,* 12**,** 7233-8.

WALTER, K., MIN, J. L., HUANG, J., CROOKS, L., MEMARI, Y., MCCARTHY, S., PERRY, J. R., XU, C., FUTEMA, M., LAWSON, D., IOTCHKOVA, V., SCHIFFELS, S., HENDRICKS, A. E., DANECEK, P., LI, R., FLOYD, J., WAIN, L. V., BARROSO, I., HUMPHRIES, S. E., HURLES, M. E., ZEGGINI, E., BARRETT, J. C., PLAGNOL, V., RICHARDS, J. B., GREENWOOD, C. M., TIMPSON, N. J., DURBIN, R. & SORANZO, N. 2015. The UK10K project identifies rare variants in health and disease. *Nature,* 526**,** 82-90.

WALTHER, A., HOULSON, R. & TOMLINSON, I. 2008. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut*.

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. & YANG, H. 2013. Preoperative platelet count associates with survival and distant metastasis in surgically resected colorectal cancer patients. *J Gastrointest Cancer,* 44**,** 293-304.

WANG, C., TAO, B., LI, S., LI, B., WANG, X., HU, G., LI, W., YU, Y., LU, Y. & LIU, J. 2014. Characterizing the role of PCDH9 in the regulation of glioma cell apoptosis and invasion. *J Mol Neurosci,* 52**,** 250-60.

WANG, H., PENG, J., WANG, B., LU, X., ZHENG, J. Z., WANG, K., TU, X. M. & FENG, C. 2017. Inconsistency Between Univariate and Multiple Logistic Regressions. *Shanghai Arch Psychiatry,* 29**,** 124-128.

WANG, J., LI, S., LIU, Y., ZHANG, C., LI, H. & LAI, B. 2020. Metastatic patterns and survival outcomes in patients with stage IV colon cancer: A population-based analysis. *Cancer Med,* 9**,** 361-373.

WANG, W. Y., BARRATT, B. J., CLAYTON, D. G. & TODD, J. A. 2005. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet,* 6**,** 109-18.

WANG, Z., LIANG, L., YU, Y., WANG, Y., ZHUANG, R., CHEN, Y., CUI, Y., ZHOU, Y. & LIU, T. 2016. Primary Tumour Resection Could Improve the Survival of Unresectable Metastatic Colorectal Cancer Patients Receiving Bevacizumab-Containing Chemotherapy. *Cell Physiol Biochem,* 39**,** 1239-46.

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. & 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**,** 631-9.

WASSERBERG, N. & KAUFMAN, H. S. 2007. Palliation of colorectal cancer. *Surgical Oncology,* 16**,** 299-310.

WATT, D. G., MARTIN, J. C., PARK, J. H., HORGAN, P. G. & MCMILLAN, D. C. 2015. Neutrophil count is the most important prognostic component of the differential white cell count in patients undergoing elective surgery for colorectal cancer. *Am J Surg,* 210**,** 24-30.

WATTS, K., WILLS, C., MADI, A., PALLES, C., MAUGHAN, T. S., KAPLAN, R., AL-TASSAN, N. A., KERR, R., KERR, D., GRAY, V., WEST, H., HOULSTON, R. S., ESCOTT-PRICE, V. & CHEADLE, J. P. 2021. Genome-wide association studies of toxicity to oxaliplatin and fluoropyrimidine chemotherapy with or without cetuximab in 1800 patients with advanced colorectal cancer. *Int J Cancer*.

WEIDEMÜLLER, P., KHOLMATOV, M., PETSALAKI, E. & ZAUGG, J. B. 2021. Transcription factors: Bridge between cell signaling and gene regulation. *Proteomics,* 21**,** e2000034.

WENZEL-SEIFERT, K. & SEIFERT, R. 2003. Functional differences between human formyl peptide receptor isoforms 26, 98, and G6. *Naunyn Schmiedebergs Arch Pharmacol,* 367**,** 509-15.

WHITE, A., IRONMONGER, L., STEELE, R. J. C., ORMISTON-SMITH, N., CRAWFORD, C. & SEIMS, A. 2018. A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK. *BMC Cancer,* 18**,** 906.

WICKHAM, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

WICKHAM, H., FRANÇOIS, R., HENRY, L. & MÜLLER, K. 2020a. dplyr: A Grammar of Data Manipulation. R package version 0.8.5.

WICKHAM, H., HESTER, J. & CHANG, W. 2020b. devtools: Tools to Make Developing R Packages Easier. R package version 2.3.2.
WIECHENS, N., SINGH, V., GKIKOPOULOS, T., SCHOFIELD, P., ROCHA, S. & OWEN-HUGHES, T. 2016. The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLoS Genet,* 12**,** e1005940.
WILSON, P. M., DANENBERG, P. V., JOHNSTON, P. G., LENZ, H. J. & LADNER, R. D. 2014. Standing the test of time: targeting thymidylate biosynthesis in cancer therapy. *Nat Rev Clin Oncol,* 11**,** 282-98.
WINSOR, C. P. 1932. The Gompertz Curve as a Growth Curve. *Proc Natl Acad Sci U S A,* 18**,** 1-8.
WOLPIN, B. M. & MAYER, R. J. 2008. Systemic treatment of colorectal cancer. *Gastroenterology,* 134**,** 1296-310.
WRIGHT, J. B., BROWN, S. J. & COLE, M. D. 2010. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol,* 30**,** 1411-20.
WRIGHT, S. P. 1992. Adjusted P-Values for Simultaneous Inference. *Biometrics,* 48**,** 1005-1013.
XUE, A., WU, Y., ZHU, Z., ZHANG, F., KEMPER, K. E., ZHENG, Z., YENGO, L., LLOYD-JONES, L. R., SIDORENKO, J., WU, Y., AGBESSI, M., AHSAN, H., ALVES, I., ANDIAPPAN, A., AWADALLA, P., BATTLE, A., BEUTNER, F., BONDER, MARC J., BOOMSMA, D., CHRISTIANSEN, M., CLARINGBOULD, A., DEELEN, P., ESKO, T., FAVÉ, M.-J., FRANKE, L., FRAYLING, T., GHARIB, S., GIBSON, G., HEMANI, G., JANSEN, R., KÄHÖNEN, M., KALNAPENKIS, A., KASELA, S., KETTUNEN, J., KIM, Y., KIRSTEN, H., KOVACS, P., KROHN, K., KRONBERG-GUZMAN, J., KUKUSHKINA, V., KUTALIK, Z., LEE, B., LEHTIMÄKI, T., LOEFFLER, M., MARIGORTA, U. M., METSPALU, A., MILANI, L., MÜLLER-NURASYID, M., NAUCK, M., NIVARD, M., PENNINX, B., PEROLA, M., PERVJAKOVA, N., PIERCE, B., POWELL, J., PROKISCH, H., PSATY, B., RAITAKARI, O., RING, S., RIPATTI, S., ROTZSCHKE, O., RUËGER, S., SAHA, A., SCHOLZ, M., SCHRAMM, K., SEPPÄLÄ, I., STUMVOLL, M., SULLIVAN, P., TEUMER, A., THIERY, J., TONG, L., TÖNJES, A., VAN DONGEN, J., VAN MEURS, J., VERLOUW, J., VÖLKER, U., VÕSA, U., YAGHOOTKAR, H., ZENG, B., MCRAE, A. F., VISSCHER, P. M., ZENG, J., YANG, J. & E, Q. C. 2018. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications,* 9**,** 2941.
YAMADA, Y., TAKAHARI, D., MATSUMOTO, H., BABA, H., NAKAMURA, M., YOSHIDA, K., YOSHIDA, M., IWAMOTO, S., SHIMADA, K., KOMATSU, Y., SASAKI, Y., SATOH, T., TAKAHASHI, K., MISHIMA, H., MURO, K., WATANABE, M., SAKATA, Y., MORITA, S., SHIMADA, Y. & SUGIHARA, K. 2013. Leucovorin, fluorouracil, and oxaliplatin plus bevacizumab versus S-1 and oxaliplatin plus bevacizumab in patients with metastatic colorectal cancer (SOFT): an open-label, non-inferiority, randomised phase 3 trial. *Lancet Oncol,* 14**,** 1278-86.
YANG, D., LAI, X., XU, F., LI, Y., JIANG, W. & MA, D. 2019a. Prognosis and clinical characteristics of colorectal cancer patients with KRAS gene mutation: a 5-year follow-up study. *Int J Clin Exp Pathol,* 12**,** 409-418.

YANG, J., GUO, X., WANG, M., MA, X., YE, X. & LIN, P. 2017a. Pre-treatment inflammatory indexes as predictors of survival and cetuximab efficacy in metastatic colorectal cancer patients with wild-type RAS. *Scientific Reports,* 7**,** 17166.

YANG, L., SUN, H., LIU, X., CHEN, J., TIAN, Z., XU, J., XIANG, B. & QIN, B. 2020. Circular RNA hsa_circ_0004277 contributes to malignant phenotype of colorectal cancer by sponging miR-512-5p to upregulate the expression of PTMA. *Journal of Cellular Physiology,* n/a.

YANG, Q., FENG, M., MA, X., LI, H. & XIE, W. 2017b. Gene expression profile comparison between colorectal cancer and adjacent normal tissues. *Oncol Lett,* 14**,** 6071-6078.

YANG, Y., ZHANG, Q., MIAO, Y.-R., YANG, J., YANG, W., YU, F., WANG, D., GUO, A.-Y. & GONG, J. 2019b. SNP2APA: a database for evaluating effects of genetic variants on alternative polyadenylation in human cancers. *Nucleic Acids Research,* 48**,** D226-D232.

YARBRO, J. W., PAGE, D. L., FIELDING, L. P., PARTRIDGE, E. E. & MURPHY, G. P. 1999. American Joint Committee on Cancer prognostic factors consensus conference. *Cancer,* 86**,** 2436-46.

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.

YASHIRO, M. 2014. Ulcerative colitis-associated colorectal cancer. *World J Gastroenterol,* 20**,** 16389-97.

YOSHIHAMA, S., ROSZIK, J., DOWNS, I., MEISSNER, T. B., VIJAYAN, S., CHAPUY, B., SIDIQ, T., SHIPP, M. A., LIZEE, G. A. & KOBAYASHI, K. S. 2016. NLRC5/MHC class I transactivator is a target for immune evasion in cancer. *Proc Natl Acad Sci U S A,* 113**,** 5999-6004.

YOSHIOKA, A., TANAKA, S., HIRAOKA, O., KOYAMA, Y., HIROTA, Y., AYUSAWA, D., SENO, T., GARRETT, C. & WATAYA, Y. 1987. Deoxyribonucleoside triphosphate imbalance. 5-Fluorodeoxyuridine-induced DNA double strand breaks in mouse FM3A cells and the mechanism of cell death. *J Biol Chem,* 262**,** 8235-41.

YURKO, R., ROEDER, K., DEVLIN, B. & G'SELL, M. 2021. H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives. *Ann Hum Genet,* 85**,** 97-100.

ZHAN, Y., SEREGIN, S. S., CHEN, J. & CHEN, G. Y. 2016. Nod1 Limits Colitis-Associated Tumorigenesis by Regulating IFN-γ Production. *J Immunol,* 196**,** 5121-9.

ZHANG, H., BEREZOV, A., WANG, Q., ZHANG, G., DREBIN, J., MURALI, R. & GREENE, M. I. 2007. ErbB receptors: from oncogenes to targeted cancer therapies. *J Clin Invest,* 117**,** 2051-8.

ZHANG, J., LEE, D., DHIMAN, V., JIANG, P., XU, J., MCGILLIVRAY, P., YANG, H., LIU, J., MEYERSON, W., CLARKE, D., GU, M., LI, S., LOU, S., XU, J., LOCHOVSKY, L., UNG, M., MA, L., YU, S., CAO, Q., HARMANCI, A., YAN, K.-K., SETHI, A., GÜRSOY, G., SCHOENBERG, M. R., ROZOWSKY, J., WARRELL, J., EMANI, P., YANG, Y. T., GALEEV, T., KONG, X., LIU, S., LI, X., KRISHNAN, J., FENG, Y., RIVERA-MULIA, J. C., ADRIAN, J., BROACH, J. R., BOLT, M., MORAN, J., FITZGERALD, D., DILEEP, V., LIU, T., MEI, S., SASAKI, T., TREVILLA-GARCIA, C., WANG, S., WANG, Y., ZANG, C., WANG, D., KLEIN, R. J., SNYDER, M., GILBERT, D. M., YIP, K., CHENG, C., YUE, F., LIU, X. S., WHITE, K. P. & GERSTEIN, M.

2020. An integrative ENCODE resource for cancer genomics. *Nature Communications,* 11**,** 3696.

ZHANG, K., CIVAN, J., MUKHERJEE, S., PATEL, F. & YANG, H. 2014a. Genetic variations in colorectal cancer risk and clinical outcome. *World J Gastroenterol,* 20**,** 4167-77.

ZHANG, L., MA, L. & ZHOU, Q. 2011. Overall and KRAS-specific results of combined cetuximab treatment and chemotherapy for metastatic colorectal cancer: a meta-analysis. *Int J Colorectal Dis,* 26**,** 1025-33.

ZHANG, M., CUI, F., LU, S., LU, H., JIANG, T., CHEN, J., ZHANG, X., JIN, Y., PENG, Z. & TANG, H. 2014b. Increased expression of prothymosin-α, independently or combined with TP53, correlates with poor prognosis in colorectal cancer. *Int J Clin Exp Pathol,* 7**,** 4867-76.

ZHAO, Z., FU, Y.-X., HEWETT-EMMETT, D. & BOERWINKLE, E. 2003. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene,* 312**,** 207-213.

ZHENG, H. F., RONG, J. J., LIU, M., HAN, F., ZHANG, X. W., RICHARDS, J. B. & WANG, L. 2015. Performance of genotype imputation for low frequency and rare variants from the 1000 genomes. *PLoS One,* 10**,** e0116487.

ZHOU, S., HAWLEY, J. R., SOARES, F., GRILLO, G., TENG, M., MADANI TONEKABONI, S. A., HUA, J. T., KRON, K. J., MAZROOEI, P., AHMED, M., ARLIDGE, C., YUN, H. Y., LIVINGSTONE, J., HUANG, V., YAMAGUCHI, T. N., ESPIRITU, S. M. G., ZHU, Y., SEVERSON, T. M., MURISON, A., CAMERON, S., ZWART, W., VAN DER KWAST, T., PUGH, T. J., FRASER, M., BOUTROS, P. C., BRISTOW, R. G., HE, H. H. & LUPIEN, M. 2020. Noncoding mutations target cis-regulatory elements of the FOXA1 plexus in prostate cancer. *Nature Communications,* 11**,** 441.

ZOCCHE, D. M., RAMIREZ, C., FONTAO, F. M., COSTA, L. D. & REDAL, M. A. 2015. Global impact of KRAS mutation patterns in FOLFOX treated metastatic colorectal cancer. *Front Genet,* 6**,** 116.

ZORI COMBA, A., BLAJMAN, C., RICHARDET, E., BELLA, S., VILANOVA, M., COPPOLA, F., VAN KOOTEN, M., RODGER, J., GIGLIO, R., BALBIANI, L., PERAZZO, F., MONTIEL, M., CHACON, M., PUJOL, F., MICKIEWICZ, E., CAZAP, E., RECONDO, G., LASTIRI, F., SIMON, J., WASSERMAN, E. & SCHMILOVICH, A. 2001. A randomised phase II study of oxaliplatin alone versus oxaliplatin combined with 5-fluorouracil and folinic acid (Mayo Clinic regimen) in previously untreated metastatic colorectal cancer patients. *Eur J Cancer,* 37**,** 1006-13.

ZWELLING, L. A., ANDERSON, T. & KOHN, K. W. 1979. DNA-protein and DNA interstrand cross-linking by cis- and trans-platinum(II) diamminedichloride in L1210 mouse leukemia cells and relation to cytotoxicity. *Cancer Res,* 39**,** 365-9.

8. Appendices

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

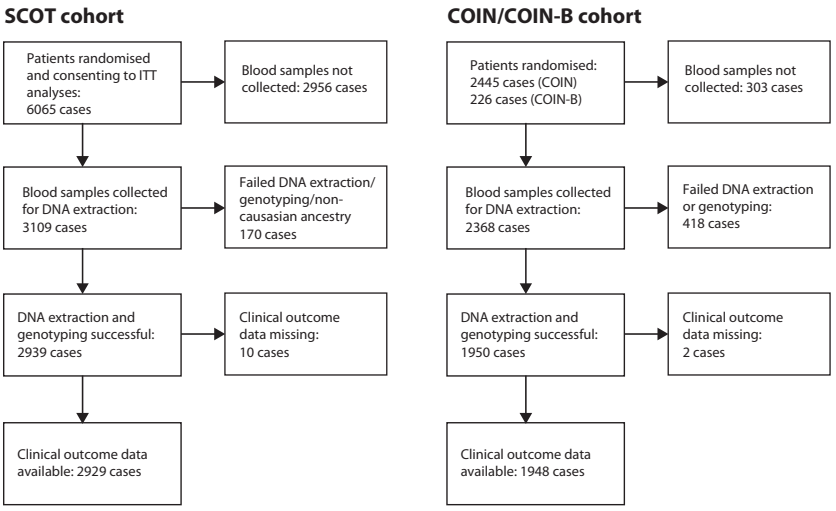

Figure 1. CONSORT diagram showing flow of patients analyzed in the study. ITT = intention to treat.

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

ARTICLE

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-β) 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.

ARTICLE

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

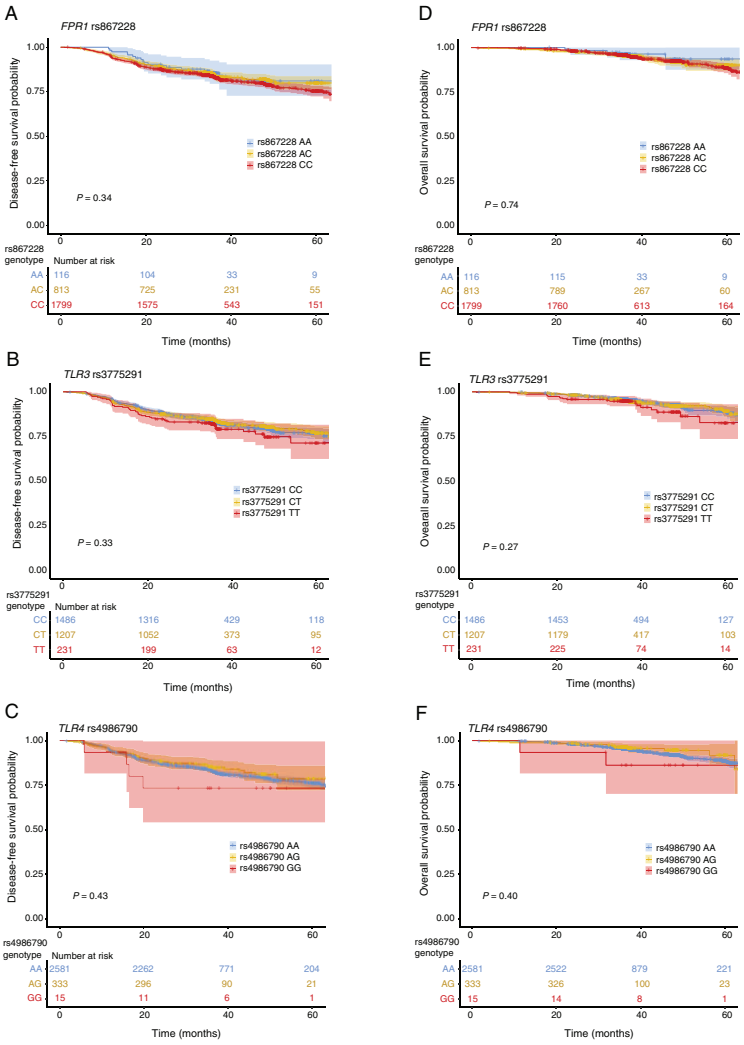

Figure 2. *FPR1*, *TLR3*, and *TLR4* loss of function (LOF) single nucleotide polymorphisms (SNPs), disease-free survival (DFS), and overall survival (OS) in Short Course in Oncology Therapy (SCOT) cohort. Kaplan Meier curves showing DFS for patients in SCOT cohort by pattern recognition receptor SNPs 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

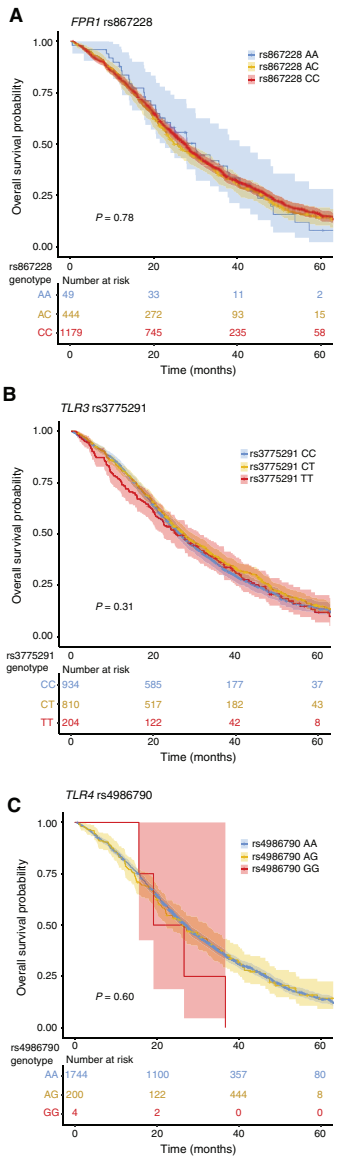

Figure 3. *FPR1*, *TLR3*, and *TLR4* loss of function (LOF) single nucleotide polymorphisms (SNPs) and overall survival (OS) in COIN/COIN-B cohort. Kaplan Meier curves showing OS for patients in combined COIN/COIN-B cohort by pattern recognition receptor SNPs rs867228 (*FPR1* c.1037A>C p.Glu346Ala) (**A**), 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

ARTICLE

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.

environmental microbiology

Environmental Microbiology (2020) **22**(6), 2150–2164 doi:10.1111/1462-2920.14977

β-Glucan is a major growth substrate for human gut bacteria related to *Coprococcus eutactus*

Anna M. Alessi,[1,2] Victoria Gray,[1,3] Freda M. Farquharson,[1] Adriana Flores-López,[1] Sophie Shaw,[3] David Stead,[1] Udo Wegmann,[2,4] Claire Shearman,[2] Mike Gasson,[2] Elaina S. R. Collie-Duguid,[3] Harry J. Flint[1] and Petra Louis[1*]

[1]*University of Aberdeen, Rowett Institute, Aberdeen, UK.*

[2]*Institute of Food Research, Norwich, UK.*

[3]*University of Aberdeen, Centre for Genome-Enabled Biology and Medicine, Aberdeen, UK.*

[4]*School of Chemistry, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK.*

Summary

A clone encoding carboxymethyl cellulase activity was isolated during functional screening of a human gut metagenomic library using *Lactococcus lactis* MG1363 as heterologous host. The insert carried a glycoside hydrolase family 9 (GH9) catalytic domain with sequence similarity to a gene from *Coprococcus eutactus* ART55/1. Genome surveys indicated a limited distribution of GH9 domains among dominant human colonic anaerobes. Genomes of *C. eutactus*-related strains harboured two GH9-encoding and four GH5-encoding genes, but the strains did not appear to degrade cellulose. Instead, they grew well on β-glucans and one of the strains also grew on galactomannan, galactan, glucomannan and starch. *Coprococcus comes* and *Coprococcus catus* strains did not harbour GH9 genes and were not able to grow on β-glucans. Gene expression and proteomic analysis of *C. eutactus* ART55/1 grown on cellobiose, β-glucan and lichenan revealed similar changes in expression in comparison to glucose. On β-glucan and lichenan only, one of the four GH5 genes was strongly upregulated. Growth on glucomannan led to a transcriptional response of many genes, in particular a strong upregulation of glycoside hydrolases involved in mannan degradation. Thus, β-glucans are a major growth substrate for species related to *C. eutactus*, with glucomannan and galactans alternative substrates for some strains.

Received 3 July, 2019; revised 27 February, 2020; accepted 3 March, 2020. *For correspondence. E-mail p.louis@abdn.ac.uk; Tel. 44(+0) 1224 438735; Fax 44(+0) 1224 438699.

Introduction

Dietary fibre originates from plant cell wall polysaccharides including cellulose, hemicellulose and pectins (Flint *et al*., 2012a). Their molecular structure is highly heterogeneous due to the presence of different monosaccharides, which are bound by a variety of glycosidic bonds. They are recalcitrant to digestion in the small intestine and reach the colon, where they serve as a substrate for microbial fermentation, which leads to the formation of short-chain fatty acids (SCFAs, mainly acetate, propionate and butyrate) and gases (Flint *et al*., 2012a). Indigestible plant storage polysaccharides and oligosaccharides are also widely regarded as belonging to the fibre-fraction of foods, as they reach the large intestine intact (Howlett *et al*., 2010; Slavin, 2013). An adequate supply of fibre and its efficient degradation is essential in fuelling the numerous health-promoting actions of the gut microbiota, in particular the provision of beneficial SCFAs (Flint *et al*., 2012b; Louis and Flint, 2017). The fermentability of hemicelluloses, consisting of different types of polysaccharides (arabinoxylans, β-glucans, mannans, xyloglucans, etc.) is relatively high in the human gut and has been reported to rely on hydrolytic action of specific isolates belonging to many genera, including *Roseburia* and *Bacteroides* (Flint *et al*., 2012a; Sheridan *et al*., 2016; Tuncil *et al*., 2017). In contrast, the breakdown of cellulose by the human gut microbiota is less efficient and appears to be restricted to few species (Cann *et al*., 2016). *Ruminococcus champanellensis* is closely related to *Ruminococcus flavefaciens*, a major cellulose degrader in the herbivore GI tract (Flint *et al*., 2008; Chassard *et al*., 2011). Cellulolytic activity was also reported for a *Bacteroides* species isolated from the human gut (Robert *et al*., 2007). There is evidence for inter-individual variation in cellulose-degrading gut microbes, which appears to correlate with whether methanogenic Archaea are present in an individual (Chassard *et al*., 2010).

The colonic microbiota is highly complex and our knowledge to date of which microbes are instrumental in fibre breakdown is incomplete. Numerous carbohydrate-active

enzymes involved in fibre breakdown have been characterized and large human gut-derived data sets are available for genomic and metagenomic mining. However, it is often difficult to deduce function from sequence alone, as many glycoside hydrolase families comprise enzymes with different substrate specificities (Lombard *et al*., 2014). Functional metagenomics, which relies on functional expression of environmental genes in a heterologous host, can be utilized to identify microbes that are involved in fibre breakdown in the gut and reveal novel enzymatic functions. This approach has successfully been applied to human gut microbiota (Tasse *et al*., 2010; Cecchini *et al*., 2013). The heterologous expression host commonly used is *Escherichia coli*, which has been reported to successfully express genes from a wide range of organisms (Handelsman, 2004). However, an in silico analysis of 32 prokaryotic genome sequences for the presence of expression signals functional in *E. coli* suggests that only approximately 40% of genes would be successfully expressed in this host, with extensive variation (7%–73%) between different organisms (Gabor *et al*., 2004). Furthermore, post-translational processes, such as protein folding, insertion into the cell membrane or secretion from the cell may also differ between different microbes. The use of alternative expression hosts for metagenomic libraries may therefore improve the recovery of novel genes from metagenomic libraries, which has been demonstrated for functional metagenomic studies from other environments (McMahon *et al*., 2012).

Here, we report the comparative analysis of *E. coli* XL1 Blue and *Lactococcus lactis* MG1363 as hosts for functional screening of a human faecal metagenomic library on a range of different dietary carbohydrates. *L. lactis* belongs to low %G + C Firmicutes, is well-characterized and widely used as an alternative host to *E. coli* for heterologous gene expression, and genetic tools and vectors are available (Pontes *et al*., 2011). This led to the identification of a clone carrying a glycoside hydrolase (GH) family 9 gene with sequence identity to *Coprococcus eutactus* ART55/1. As this GH family is usually associated with cellulose breakdown, we investigated the breakdown of beta-linked glucans in this strain and its gene and protein expression response to growth on different substrates.

Results

Screening for glycoside hydrolase activities from a metagenomic library in Escherichia coli *and* Lactococcus lactis

A human gut microbiome metagenomic library (6146 clones, average insert size estimate 2.5 kb), constructed in shuttle vector pTRKL2 and transformed into *E. coli* XL1 Blue, was functionally screened for glycoside hydrolase (GH) activities using seven carbohydrate substrates. Enzyme activity was confirmed in 16 clones after re-streaking on the respective media. Positive clones were found for all substrates apart from polygalacturonic acid and rhamnopyranoside, with the

Table 1. Metagenomic clones with glycoside hydrolase activity detected by functional screening of clone libraries generated in *E. coli* XL1 Blue and *L. lactis* MG1363.

			Enzyme activity[a]						
Clone ID	Length (nt)	Host	AF	S	CMC	L	X	Match organism	ORF matches with relevance for carbohydrate metabolism[b]
P3H22	3203	*E. coli*	+					*Bacteroides plebeius*	Glycoside hydrolase, alpha-L-arabinofuranosidase, beta-galactosidase
P3B15	3470	*E. coli*	+++					*Eubacterium rectale*	Alpha-L-arabinofuranosidase, alpha-N-arabinofuranosidase
P1P9	5693	*E. coli*		+				*Bacteroides plebeius*	SusD family nutrient uptake, carbohydrate-binding protein
P5H21	4130	*E. coli*		+				*Bacteroides uniformis*	Type I pullulanase
P5E1	3949	*E. coli*		+	+/–			*Bacteroides* sp.	6-Phosphogluconate dehydrogenase
P5D24	5531	*E. coli*		+	+/–			*Alistipes putredinis*	
P3N11	3927	*E. coli*		+/–	+/–			Bacteroidales	
P1E14	3637	*E. coli*		+				*Eubacterium* sp.	Alpha amylase
P1I16	2754	*E. coli*		++	+			*Faecalibacterium* sp.	
P3C3	2202	*E. coli*		++	+			Lachnospiraceae	Polysaccharide deacetylase
P1D18	2918	*E. coli*		+	+/–			*Anaerostipes hadrus*	
P8I17	4094	*E. coli*		++	+			*Anaerostipes hadrus*	Glucosamine-6-phosphate deaminase, cell wall hydrolase
P5J3	4246	*E. coli*		++	+			Firmicutes	
P7I21	5402	*E. coli*		+				*Coprobacillus* sp.	
P5M23	2994	*E. coli*		+	+/–				
P9N7	2655	*E. coli*			+++	+++	++	*Alistipes senegalensis*	
P20A8	2643	*L. lactis*			+++	+++	++	*Coprococcus eutactus*	Bacterial surface proteins containing Ig-like domains

a. Substrates used for screening: AF, 4-methylumbelliferyl α-L-arabinofuranoside; S, potato starch; CMC, carboxymethyl cellulose; L, lichenan; X, oat spelt xylan. No clones with activity on polygalacturonic acid or 4-methylumbelliferyl α-L-rhamnopyranoside were detected. Level of activity detected is based on visual inspection of clearing zones on substrate-containing plates.
b. Full details of blast results of clone sequences are given in the Supporting Information Table S1.

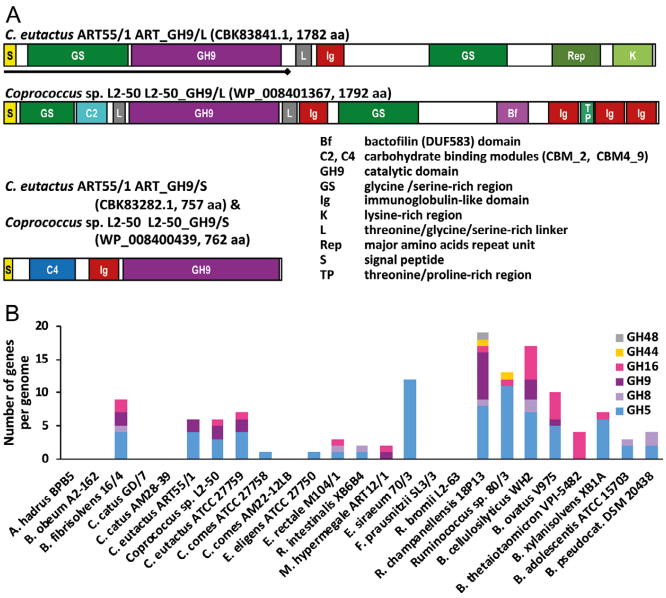

Fig. 1. Glycoside hydrolase genes associated with β-glucanase activity in human gut bacteria.
A. GH9-domain containing genes (short genes designated GH9/S and long genes GH9/L) in *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50. Accession numbers and deduced length in amino acids (aa) are given in brackets. The start codon for L2-50_GH9/L was re-assigned to position 18 of WP_008401367.1 based on the presence of a ribosome-binding site (GGAAG, eight nucleotides upstream) and a signal peptide motif. The line below ART_GH9/L indicates the region covered by clone P20A8 from the metagenomic library. Domain structure predictions are based on PFAM, PROSITE, InterPro and SMART databases searching.
B. Genome carriage of glycoside hydrolase (GH) gene families associated with β-glucanase activity in human gut bacteria.

highest number of clones on starch and carboxymethyl cellulose (Table 1). The comparative analysis of most sequenced inserts from positive clones showed a high level of identity to sequences from a variety of Bacteroidetes and Firmicutes bacteria of gut origin (Table 1 and Supporting Information Table S1). Sequence analysis of open reading frames revealed homology with enzymes with the expected substrate specificity for some clones (e.g. P3H22 and P3B15, detected on α-L-arabinofuranoside; P5H21 and P1E14, detected on starch) whereas other clone sequences harboured less well-characterized open reading frames (Table 1 and Supporting Information Table S1).

The *E. coli* XL1 Blue library was pooled and transferred into *L. lactis* MG1363 (4608 clones, insert frequency estimate 75%, average insert size estimate 2.5 kb) and functionally screened on all seven carbohydrate substrates, which resulted in a total of three positive clones on the β-linked carbohydrates carboxymethyl cellulose (CMC), lichenan and xylan. Sequencing analysis revealed that all three clones contained identical insert sequences but were different from the single positive clone found on the same substrates in *E. coli* XL1 Blue (Table 1 and Supporting Information Table S1). One of the clones (P20A8) was transferred into *E. coli* XL1 Blue, but showed only very weak enzyme activity in this host. The 16 positive clones detected in *E. coli* XL1 Blue were also transformed into *L. lactis* MG1363, but none of them displayed enzyme activity in this host (data not shown). Clone P20A8 showed the presence of a truncated ORF with 100% sequence identity to *C. eutactus* ART55/1 gene CBK83841.1 (Fig. 1A). It contained a GH9 catalytic domain, suggesting that it encodes a β-glucanase.

Distribution of β-glucanase gene families among human colonic bacteria

Following detection of the GH9 catalytic domain containing clone P20A8 from *C. eutactus* ART55/1, we performed in silico analysis of β-glucanase gene families among human colonic bacteria. Glycoside hydrolases that break down β-(1,4) linkages in glucan chains belong to multiple GH families, notably GH5, GH8, GH9, GH16, GH44 and GH48 (CAZy database at www.cazy.org; Lombard *et al*., 2014). Figure 1B shows the distribution of the best characterized β-glucanase gene families across genomes available for selected human colonic bacteria from CAZy spanning Firmicutes, Bacteroidetes and Actinobacteria. Only GH5 is widely distributed (16/25 genomes) and this family is known to include enzymes with a very diverse range of specificities. In contrast, GH48 (generally encoding cellobiohydrolases) occurs only in the cellulolytic species *R. champanellensis* while GH44 (which has been implicated in xyloglucan utilization) is limited to two

species of *Ruminococcus*. The GH9 family, which includes many bacterial cellulases among cellulolytic bacteria from gut and non-gut habitats, also shows a limited distribution among human colonic anaerobes (Fig. 1B). This prompted us to investigate the characteristics of *Coprococcus*-related isolates and their GH9 genes further.

GH9 genes of Coprococcus-related gut isolates

Genome analysis of *C. eutactus* ART55/1 and a related strain, *Coprococcus* sp. L2-50 (96% identity of the respective 16S rRNA gene sequences AY350746 and AJ270491; Fig. 2), revealed that each organism harboured two genes containing GH9 catalytic domains (Fig. 1A). All four GH9 enzymes contained a typical N-terminal lipoprotein signal peptide and showed the presence of two conserved aspartate residues and a glutamate residue involved in the catalytic reaction (Kurokawa *et al*., 2002; Pereira *et al*., 2009; Pereira *et al*., 2010). The shorter genes ART_GH9/S (CBK83282.1) and L2-50_GH9/S (WP_008400439.1) share a high level of identity (60%) and similar multi-domain structure. The catalytic domains of the longer proteins ART_GH9/L (CBK83841.1) and L2-50_GH9/L (WP_008401367.1) share 48% identity and the complete proteins exhibited a more diverse multi-domain architecture that differed between both enzymes (Fig. 1A).

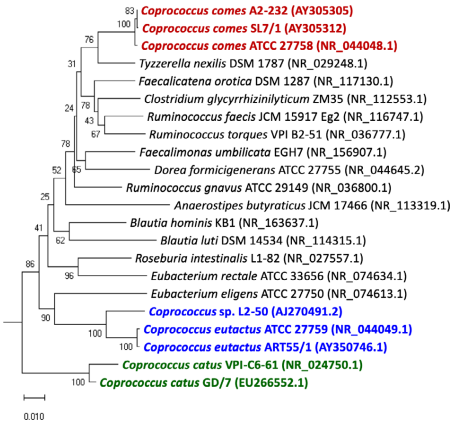

Fig. 2. Evolutionary relationships of *Coprococcus* species and their closest human gut bacteria relatives, inferred using the Neighbour-Joining method (Saitou and Nei, 1987). Bootstrap values (Felsenstein, 1985) from 500 replications are shown at branches and evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al*., 2004) using Mega X (Kumar *et al*., 2018).

Phylogenetic analysis, based on multiple amino acid alignments of the catalytic domains, showed that ART_GH9/S and L2-50_GH9/S have >50% sequence identity to GH9 enzymes from other Clostridiales species including Lachnospiraceae bacteria, *Butyrivibrio* and *Ruminococcus* species within theme D of GH9 cellulases (Supporting Information Fig. S1). All of these enzymes display modular architecture similar to ART_GH9/S and L2-50_GH9/S with the presence of a carbohydrate-binding domain (CBM4_9, pfam02018), Ig-like domain (cd02850) and glycosyl hydrolase family 9 catalytic domain (pfam00759). The amino acid sequences of the GH9/L catalytic modules from both strains showed a lower level of identity (40%–50%) to other Clostridiales species belonging to Lachnospiraceae and *Ruminococcus* (Supporting Information Fig. S1).

The genus *Coprococcus* within the Lachnospiraceae family of Firmicutes contains three species, however, they are not phylogenetically closely related (Fig. 2). Sequence similarity searches of both GH9 genes identified in *C. eutactus* ART55/1 against reference genomes of *C. eutactus*, *Coprococcus catus* and *Coprococcus comes* revealed that both genes were present in eight *C. eutactus* strains including the type strain *C. eutactus* ATCC 27759 (ART_GH9/S, CBK83282.1, query coverage >91%, sequence identity >75%; ART_GH9/L, CBK83841.1, query coverage >97%, sequence identity >68%). No significant similarity was found with either GH9 gene in the *C. catus* and *C. comes* genomes. CAZyme analysis of several genomes of different *Coprococcus* species further confirmed that the GH9 genes were only present in *C. eutactus*, whereas in the other species, no β-glucanase-related genes were found apart from a single GH5 gene in one of the examined *C. comes* strains (Fig. 1B).

Catalytic activities of the GH family 9 enzymes from C. eutactus *ART55/1 and Coprococcus sp. L2-50*

The full length GH9 genes including predicted promoter sites were cloned in the same orientation into shuttle vector pTRKL2 and transformed into *E. coli* XL1 Blue and *L. lactis* MG1363. The transformants exhibited enzyme activity on substrate-containing plates using the Congo Red detection method (Fig. 3A). The enzymes from *Coprococcus* sp. L2-50 were functionally expressed in both hosts and showed activity on CMC- and lichenan-containing plates. In contrast, the enzymes from *C. eutactus* ART55/1 were functionally expressed only in *L. lactis* MG1363. The *E. coli* XL1 Blue transformants were devoid of lichenase activity and showed only limited activity on CMC-containing plates (Fig. 3A). Analysis of codon usage of the two GH9 genes from *C. eutactus* ART55/1 revealed that differences in codon usage may be

responsible for their poor expression in *E. coli* XL1 Blue, as two codons of infrequent use in *E. coli* (Zhang *et al*., 1991) were present at relatively high frequency (ART_GH9/S AGA 14.5/1000, AUA 25.1/1000; ART_GH9/L AGA 9.5/1000, AUA 26.9/1000). The *Coprococcus* sp. L2-50 genes, on the other hand, had mostly lower frequencies (L2-50_GH9/S AGA 11.8/1000, AUA 0/1000; L2-50_GH9/L AGA 6.6/1000, AUA 7.7/1000) in line with their higher expression levels in *E. coli* XL1 Blue. The enzyme catalytic activities were also examined in supernatants and cell-free extracts of grown cultures using a reducing sugar assay, which agreed with the plate assay results. Enzyme activities for *E. coli* XL1 Blue transformants were mainly associated with the cell-free extracts, whereas activities from *L. lactis* MG1363 cultures were mainly detected in the supernatant fraction (Supporting Information Table S2). Thus, the cloned enzymes appear to be secreted by *L. lactis* MG1363.

To investigate the functionality of the different domains present in the shorter GH9 gene that was strongly conserved between both strains (Fig. 1A), different constructs that lacked either the carbohydrate-binding domain and/or the Ig-like domain were generated of L2-50_GH9/S (Fig. 3B). The constructs were cloned and overexpressed in *E. coli* BL21 (DE3), and enzyme activities on various substrates were determined. Enzyme activities of the construct lacking the CBM did not differ from the complete enzyme, but the deletion of the Ig-like domain abolished catalytic activity (Fig. 3B). The highest activity was found on β-glucan, followed by lichenan.

Growth of Coprococcus species on different carbohydrates

We examined the ability of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 to utilize a wide range of carbohydrate substrates by measuring optical density and a drop in pH. *Coprococcus* sp. L2-50 exhibited growth on glucose, cellobiose, β-glucan and lichenan as well as very limited growth on potato starch, but no utilization could be detected on laminarin, glucomannan, galactomannan, mannan, galactan, xylan, xyloglucan, arabinoxylan and pullulan (Fig. 4A and Supporting Information Fig. S2). In addition to the substrates utilized by *Coprococcus* sp. L2-50, *C. eutactus* ART55/1 was also able to grow on glucomannan, galactomannan, galactan and potato starch and a limited pH drop was also detected on mannan (Fig. 4 and Supporting Information Fig. S2). Medium pH was tracked for up to 11 days on the insoluble substrates Sigmacell type 50, acid-swollen cellulose and filter paper for both *C. eutactus*-related strains. No decrease in pH was seen (data not shown), indicating a failure to utilize these substrates. Thus, both strains are non-cellulolytic but are able to utilize certain soluble β-glucans, suggesting a possible role for the GH9 enzymes given their activity against these substrates (Fig. 3). Substrate utilization

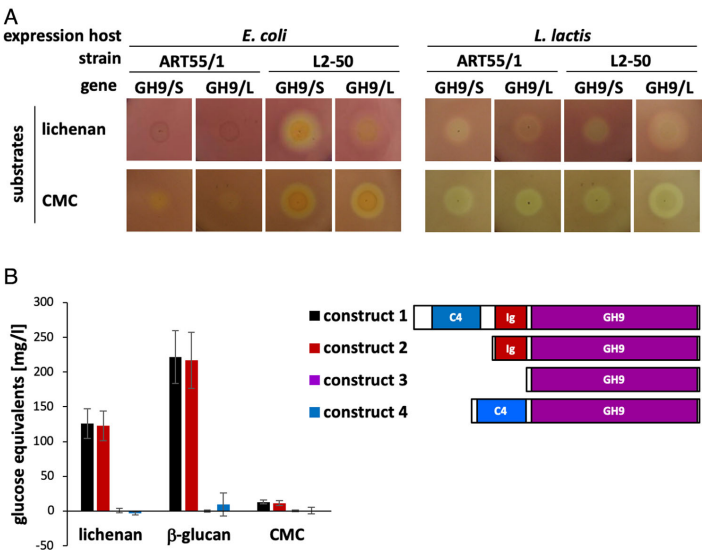

Fig. 3. Enzyme activities of GH9 genes from *Coprococcus*-related strains.
A. Fibrolytic activity of *E. coli* and *L. lactis* transformants with *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 GH9 genes on lichenan- or CMC-containing agar plates after overnight incubation of 10 μl of a freshly grown overnight culture and staining with Congo Red.
B. Enzyme activity of different constructs of *Coprococcus* sp. L2-50 L2-50_GH9/S missing either the carbohydrate binding and/or Ig-like domain (for domain designations see Fig. 1). Constructs were cloned and overexpressed in *E. coli* BL21 (DE3). Release of reducing sugars from lichenan, β-glucan and CMC was determined by Lever assay over 1 h of incubation and background activity of *E. coli* BL21 (DE3) carrying the empty vector were subtracted. Mean and standard deviation of three replicates.

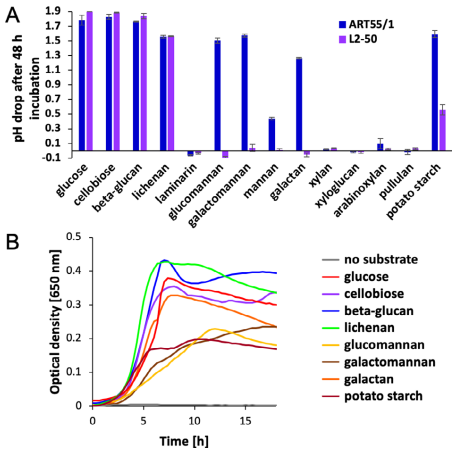

Fig. 4. Carbohydrate utilization of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 on different substrates.
A. Medium pH drop after growth in 96-well plates after 48 h of incubation (mean and standard deviation of triplicate cultures). Growth on cellulose was carried out in Hungate tubes to enable longer incubation periods, but the pH did not change after 11 days of incubation.
B. Representative growth curves of *C. eutactus* ART55/1 during growth in 96-well plates. No or very little growth was observed on laminarin, xylan, xyloglucan, arabinoxylan and pullulan (OD <0.07).

was also examined of further strains from the three different *Coprococcus* species, which revealed that *C. eutactus* ATCC 27759 showed a very similar behaviour to *C. eutactus* ART55/1, but the three *C. comes* strains showed good growth only on glucose and *C. catus* GD/7 only showed very limited growth on potato starch (Supporting Information Fig. S2). This is in agreement with both the phylogenetic placement of the species as well as with the absence of the respective glycoside hydrolase genes in those strains.

Induction of β-glucanase activity during growth on β-glucan and transcriptional response to growth on different carbohydrates in C. eutactus *ART55/1*

Coprococcus eutactus ART55/1 and *Coprococcus* sp. L2-50 were grown to early stationary phase on either glucose, cellobiose or β-glucan to investigate whether β-glucanase activity was inducible in these strains. β-glucanase activity was much higher for cell extracts and supernatants from cultures grown on β-glucan than on glucose or cellobiose, especially for *C. eutactus* ART55/1 (Supporting Information Table S3). To identify differentially expressed genes (DEGs) during growth on β-glucan-type carbohydrates, *C. eutactus* ART55/1 grown on glucose was inoculated into medium containing glucose, cellobiose, β-glucan, lichenan or glucomannan and grown to exponential phase (Supporting Information Table S4) for RNA extraction and transcriptomic analysis by RNA sequencing. Principal component analysis (PCA) of transcript changes showed that *C. eutactus* ART55/1 cultures grown on glucomannan clustered most distinctly, whereas more similarity was seen between cellobiose, β-glucan and lichenan incubations (Fig. 5A). Significant DEGs [false discovery rate (FDR) <0.05 and log-fold change (LogFC) >1] compared to glucose as baseline were determined for all carbon sources. Glucomannan showed 197 DEGs compared to glucose specific to this substrate, whereas the other three substrates each only exhibited between two and seven specific DEGs (Fig. 5B). For all DEGs per substrate (including those shared with other substrates), on glucomannan almost half of the genes were upregulated (44%) relative to glucose, whereas on the other three substrates almost three quarters of the genes were downregulated (cellobiose 71%, β-glucan 72%, lichenan 68%, Fig. 5B). This trend was also reflected in the magnitude of the response (maximum fold change of upregulated DEGs 3.6, 20.7, 16.7 and 560.7 on cellobiose, β-glucan, lichenan and glucomannan, respectively, downregulated DEGs 273.8, 295.0, 231.9 and 40.6, respectively, Supporting Information Table S5).

Thirty-eight genes were significant across all four comparisons, with an additional 25 shared between β-glucan, cellobiose and lichenan (Fig. 5B). Interestingly, the complete data set of all 299 DEGs showed a very strong positive correlation for all pairwise comparisons between cellobiose, β-glucan and lichenan ($P < 0.0001$) but not for comparisons with glucomannan (Fig. 5C and Supporting Information Fig. S3). Many of the DEGs significant on all four substrates responded in the opposite direction on glucomannan relative to each of the other substrates (Fig. 5C and Supporting Information Fig. S3, data labelled in black). Gene ontology (GO) enrichment analysis identified several GO terms that were significantly over represented in the DEGs. These were mostly related to localization, transport, carbohydrate metabolic processes and hydrolase activity (Supporting Information Table S6**)**. Three of the four genes strongly downregulated on all four substrates code for a CUT1 family ABC transporter with the fourth gene encoding a GH77 (4-α-glucanotransferase). On β-glucan, lichenan and to a lesser degree cellobiose, another ABC transporter belonging to the CUT2 family was also significantly downregulated. On glucomannan, on the other hand, an ABC-transporter annotated as a multidrug transport system was strongly upregulated (Supporting Information Table S5).

The genome of *C. eutactus* ART55/1 contains 45 glycoside hydrolase genes according to the CAZy database (Supporting Information Table S7), 19 of which showed a

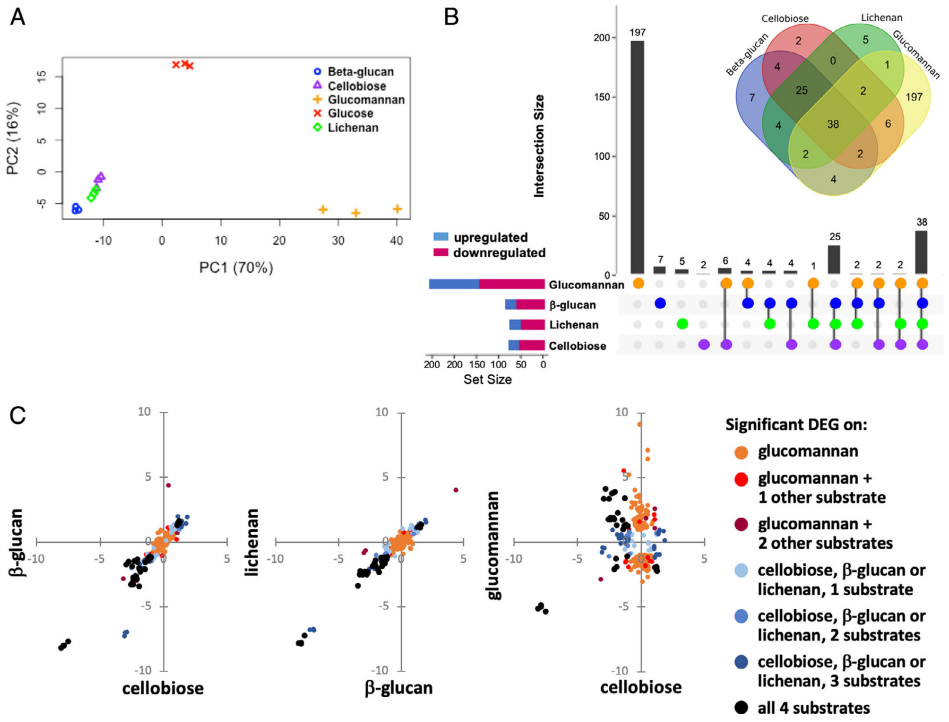

Fig. 5. Gene expression changes after growth of *C. eutactus* ART55/1 on five different substrates.
A. PCA showing separation of the samples based on differences in gene expression with samples coloured by carbon source.
B. UpSet plot and Venn diagram summarizing total number of genes differentially expressed compared to glucose, direction of change and intersections between sets of differentially expressed genes.
C. Relationship of all differentially expressed genes (DEGs, 299 genes in total) between different growth substrates, expressed as logFC. The remaining comparisons are shown in Fig. S3.

significant change in gene expression (Fig. 6). One of the four GH5 genes present in the genome was strongly upregulated on β-glucan and lichenan and to a lesser degree on glucomannan. One of the two GH9 genes (GH9/L) was significantly upregulated on all four substrates, with glucomannan showing the strongest response. For most other GH genes, the response was opposite for glucomannan compared to the other substrates, with the vast majority being upregulated on glucomannan. The strongest upregulation was observed for all five genes assigned to GH families predominantly involved in mannan degradation (GH26, GH113 and GH130; Fig. 6). All ribosomal proteins were significantly reduced (FDR <0.05) on glucomannan and several reached a logFC of over −1, whereas none of the other substrates showed any significant changes relative to glucose (Supporting Information Table S5). This is in agreement with the lower growth rate observed on glucomannan (Supporting Information Table S4).

Proteomic response to growth on different carbohydrates in C. eutactus *ART55/1*

Bacterial cells from the cultures used for gene expression analysis were also subjected to proteomics analysis by mass spectrometry. In total, 891 *C. eutactus* ART55/1 proteins were identified. Label-free quantification (LFQ) values were analysed for differential abundance with a linear model identical to the gene expression analysis with glucose as the baseline and all proteins with adjusted significance values below 0.05 were regarded as differentially abundant. PCA revealed strikingly similar relationships between the different samples as was found

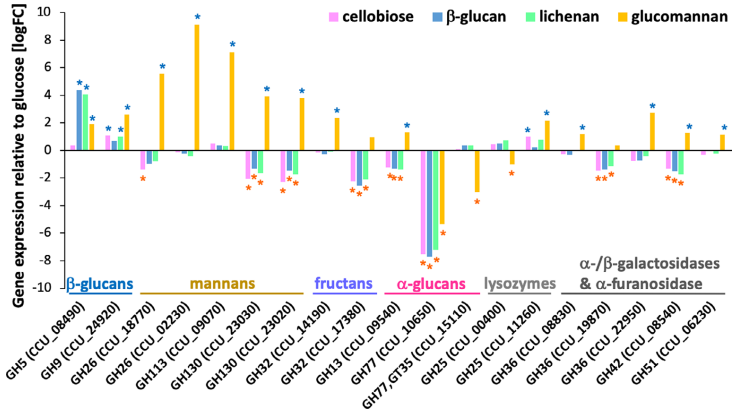

Fig. 6. Carbohydrate-active genes differentially expressed compared to glucose in *C. eutactus* ART55/1 during growth on cellobiose, β-glucan, lichenan and glucomannan.
Putative enzyme substrates based on Cazypedia (CAZypedia Consortium, 2018) are indicated above the GH genes. Significant up- and down-regulation for each of the four growth substrates relative to glucose is indicated by blue and orange asterisks, respectively.

for the gene expression data (Fig. 7A). A comparison of LFQ values with reads per kilobase per million mapped reads (RPKM) values from the gene expression analysis showed a strong positive trend between the two datasets (Fig. 7B) despite the fact that they showed only a partial overlap in significantly differentially expressed/abundant genes and proteins (Supporting Information Fig. S4), likely reflecting post-transcriptional networks regulating protein expression (Vogel and Marcotte, 2012). As for the gene expression data, glucomannan resulted in the largest difference in the proteome compared to glucose, and the overall distribution of shared proteins between different substrates was similar as well (Fig. 7C). Several expression changes identified at the transcript level were confirmed at the proteome level, including the upregulated GH5 and some of the proteins comprising the downregulated CUT1 and CUT2 ABC transporters on β-glucan-type substrates, as well as some of the carbohydrate active enzymes upregulated on glucomannan (Supporting Information Table S5).

Discussion

Functional metagenomic screening was employed here to identify genes involved in carbohydrate breakdown from human faecal microbiota. A comparison of the suitability of two cloning hosts, *E. coli* XL1 Blue and *L. lactis* MG1363, revealed large differences in their functional expression of heterologous genes. Several active clones from a wide range of bacteria within both the Bacteroidetes and Firmicutes were recovered from the *E. coli* XL1 Blue library, whereas only a single clone with β-glucanase activity was recovered after electroporation of the *E. coli* XL1 Blue library into *L. lactis* MG1363. Thus, *L. lactis* MG1363 appears to be quite limited in its ability to successfully express genes from other organisms, and alternative Gram-positive hosts (e.g. *Bacillus subtilis*, Dobrijevic *et al*., 2013) may be more suitable for future functional metagenomic studies. This study does, however, also show the limitations in using *E. coli* as heterologous host, as the clone expressing very high activity in *L. lactis* MG1363 (P20A8) did not exhibit good activity in *E. coli* XL1 Blue. *E. coli* has been estimated to express approximately 40% of enzymatic activities from diverse microbial origins based on the analysis of expression signals in microbial genomes and is strongly biased towards genes from certain groups of organisms (Uchiyama and Miyazaki, 2009). Differences in codon usage between *C. eutactus* ART55/1 and *E. coli* XL1 Blue likely contribute to the poor expression of the two GH9 genes in *E. coli*. Protein export from the cell can also significantly affect successful heterologous expression (Freudl, 2018). Fractionation of grown cultures and enzyme activity measurements showed that *E. coli* XL1 Blue was not able to export the extracellular enzymes efficiently, whereas the activity was mainly detected in the culture supernatant in *L. lactis* MG1363. Differences in signal peptide recognition, secretion mechanism, the presence of chaperones and architecture of the cell envelope can affect secretion abilities between different organisms (Mingardon *et al*., 2011; Burdette *et al*., 2018).

The glycoside hydrolase family 9 mainly consists of cellulases (Wilson and Urbanowicz, 2019) and GH9 genes

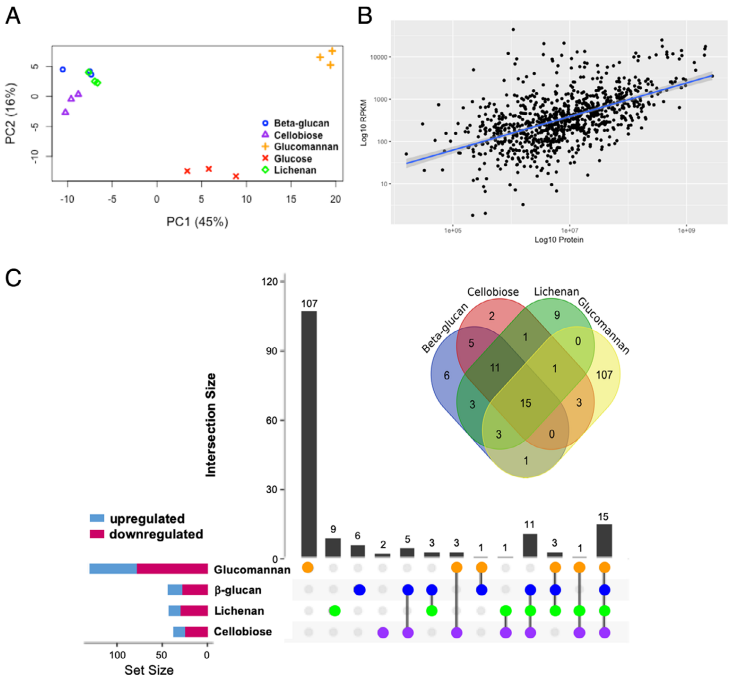

Fig. 7. Proteomic changes after growth of *C. eutactus* ART55/1 on five different substrates.
A. PCA showing separation of the samples based on differences in protein expression with samples coloured by carbon source.
B. Correlation between transcript and proteome expression levels for all detected proteins (log10 average protein expression versus log 10 average RPKM for each gene).
C. UpSet plot and Venn diagram summarizing total number of proteins differentially expressed compared to glucose, direction of change and intersections between sets of differentially expressed genes.

in the human cellulose degrader *R. champanellensis* have been shown to be involved in cellulose breakdown (Morais *et al*., 2016). This prompted us to investigate whether the *C. eutactus*-related strains ART55/1 and L2-50 are able to grow on cellulose, but this was not the case. Instead, both strains showed excellent growth on barley β-glucan and lichenan. The carriage of GH9 genes is therefore not a clear indicator of cellulose-degrading capacity in human gut bacteria. Based on sequence similarity and domain structure, ART_GH9/S and L2-50_GH9/S belong to GH9 theme D. Enzymes belonging to this group have been shown to initially cleave cellulose in a random mode and then act mainly as cellobiohydrolases (Devillard *et al*., 2004). The catalytic domains of ART_GH9/L and L2-50_GH9/L show sequence similarity with *R. champanellensis* GH9A, which was classed as an endoglucanase (Morais *et al*., 2016). More generally, GH9 enzymes have been associated with eight different catalytic specificities (EC numbers) however and these include activity against mixed-linkage beta glucans, as detected here (CAZY website www.cazy.org, Lombard *et al*. 2014). *Coprococcus* sp. L2-50 did not exhibit good growth on any of the other eight polysaccharides tested, but *C. eutactus* ART55/1 and ATCC 27759 showed some growth on glucomannan, galactomannan, galactan and starch. Genome analysis revealed that *Coprococcus* sp. L2-50 contains significantly fewer glycoside hydrolase genes [30 based on database dbCAN2 (Zhang *et al*., 2018)] than *C. eutactus* ART55/1 (43 based on dbCAN2, 45 based on CAZy, www.cazy.org) (Supporting Information Table S7). In agreement with the differences in carbohydrate degradation capacity seen for the two strains (Fig. 4), no genes belonging to GH26, GH113 or GH130, that were highly upregulated in *C. eutactus* ART55/1 on glucomannan, were identified in *Coprococcus* sp. L2-50 (Supporting Information Table S7). Despite the inability of *Coprococcus* sp. L2-50 to degrade galactan, the genome carriage of putative β-galactanases (CAZypedia Consortium, 2018) was similar between the two strains (one GH16 in *Coprococcus*

sp. L2-50 and one GH53 in *C. eutactus* ART55/1, Supporting Information Table S7). *Coprococcus* sp. L2-50, however, harboured only a single potential β-galactosidase (belonging to GH2), whereas *C. eutactus* ART55/1 encoded four (two GH1, one GH2 and one GH42). GH42 enzymes have been hypothesized to be involved in plant cell wall degradation and may work in cooperation with GH53 galactanases (Moracci, 2019).

Gene expression and proteomic analysis of *C. eutactus* ART55/1 on β-glucan-type substrates compared to glucose revealed a significant increase in one of the two GH9 genes, which was strongest on glucomannan. GH5 enzymes also hydrolyse β-glucan-type linkages and four GH5 genes are present in this strain. One of these GH5 genes was found here to be strongly upregulated, in particular on β-glucan and lichenan, making it a strong candidate for involvement in their degradation. The bioinformatic analysis showed that this GH5 protein (CCU_08490) belongs to subfamily 37, which are intracellular enzymes of bacterial origin with endo-β-1,-3/4-glycanase (EC 3.2.1.4 and EC 3.2.1.73) and cellodextrinase (EC 3.2.1.74) activities (Aspeborg *et al*., 2012). Interestingly, most GH enzymes in both strains appear not to be secreted via typical secretion systems, as they contain no predicted signal peptides (signal peptides detected in four of 30 enzymes in *Coprococcus* sp. L2-50 and seven of 45 enzymes in *C. eutactus* ART55/1, Supporting Information Table S7). The absence of predicted signal peptides on the majority of GH enzymes was previously observed in other members of the Lachnospiraceae family (Sheridan *et al*., 2016). It remains to be established whether some of the GH enzymes without predicted signal peptides are membrane-associated or contain atypical secretory signal peptides (Gagic *et al*., 2016); however, the main ecological niche of the *C. eutactus*-related strains appears to be in the breakdown of soluble and shorter length β-glucans rather than complex insoluble fibre.

The relative expression levels during growth on glucose (percentage of specific gene relative to all genes, data not shown) did not reveal a big difference between the six GH5 and GH9 genes (0.011%–0.022% for the four GH5 genes (with the lowest one upregulated), 0.026 and 0.037 for the two GH9 genes (with the higher one upregulated), overall range of all genes 0.000035%–2.55%). Therefore, differences in responses to growth on polysaccharides are likely not due to differences in basal gene expression between the GH5 and GH9 genes. The glycoside hydrolases strongly upregulated on glucomannan (five genes belonging to GH26, GH113 and GH130, see Fig. 6) on the other hand had a lower basal gene expression on glucose (0.0006%–0.0093%). In general, the overall gene expression profile was very similar on cellobiose, β-glucan and lichenan and the response in comparison to glucose tended towards downregulation, with particularly strong responses for transporters likely involved in glucose transport. Thus β-glucans may be the preferred substrates for this organism, and oligosaccharides may be transported into the cell rather than glucose during growth on these substrates. The strong upregulation of genes involved in glucomannan degradation on this substrate, on the other hand, indicates that glucomannan constitutes an alternative energy source for *C. eutactus* ART55/1. *Coprococcus* sp. L2-50, a close relative of *C. eutactus* ART55/1, did not show good growth on any of the non-β-glucan-type substrates tested here, suggesting that bacteria related to *C. eutactus* may be adapted to thrive on β-glucans in the human large intestine. The negative correlation of GH77 on β-glucan, cellobiose, lichenan and particularly glucomannan in our study suggests that availability of these substrates repressed 4-α-glucanotransferase activity in *C. eutactus* ART55/1 associated with starch utilization (Ze *et al*., 2015). The genus *Coprococcus* was increased on a high resistant starch diet in pigs and its abundance was positively correlated with starch breakdown products (Sun *et al*., 2016). Both *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 carry several GH13 genes (Supporting Information Table S7), encoding α-amylases (Ze *et al*., 2015), however, reasonably good growth on potato starch was only found for *C. eutactus* ART55/1 (and also for the second *C. eutactus* strain ATCC 27759), but final optical densities were low compared to β-glucans (Supporting Information Fig. S2). The three *C. comes* strains and *C. catus* GD/7, on the other hand, did not show good growth on any of the polysaccharides examined here and thus occupy different ecological niches from *C. eutactus*. They may degrade polysaccharides not included here or may cross-feed from breakdown products of primary polysaccharide degraders. *C. catus* grows well on fructose, but also grows on lactate, thus it is able to cross-feed on fermentation products from other bacteria (Reichardt *et al*., 2014). Both species will have to be assigned new genus names based on their phylogenetic placement as well as physiological characteristics, as *C. eutactus* is the type species of the genus *Coprococcus* (Holdeman and Moore, 1974). The evidence provided here should also be taken into consideration for the interpretation of sequence-based studies, which often do not resolve data beyond genus level. Sequence-based studies that find a change in *Coprococcus* spp. should ideally be followed up with species-specific methods such as qPCR (Reichardt *et al*., 2018).

Stimulation of *C. eutactus* may have beneficial effects on human health, as it contributes to the production of the health-promoting metabolite butyrate (Louis *et al*., 2004). *Coprococcus* spp. were also consistently associated with higher quality of life across several cohorts and depleted in depression (Valles-Colomer *et al*., 2019) and

C. eutactus showed an increase in abundance with a decrease in atopic dermatitis in an infant cohort (Nylund *et al*., 2015). Our study will aid in the development of nutritional strategies to stimulate these potentially beneficial microbes in the gut.

Experimental procedures

Bacterial strains, plasmids and growth conditions

For the metagenomic library construction, functional screening and gene cloning *Escherichia coli* XL1 Blue (Stratagene, Milton Keynes, UK) and *Lactococcus lactis* MG1363 (Gasson, 1983) were used. *E. coli* XL1 Blue clones containing plasmid pTRKL2 (O'Sullivan and Klaenhammer, 1993) were selected on BHI (#CM1136; Oxoid, Basingstoke, UK) agar plates supplemented with 150 $\mu g.ml^{-1}$ erythromycin. *L. lactis* MG1363 clones containing plasmid pTRKL2 were selected on GM17 (#CM0817; Oxoid, supplemented with 0.5% (w/v) glucose) agar plates supplemented with 5 $\mu g.ml^{-1}$ erythromycin. *E. coli* XL1 was grown at 37°C with shaking at 220 rpm, and *L. lactis* MG1363 was grown at 30°C under static conditions.

For overexpression of synthetic gene constructs of *Coprococcus* sp. L2-50 gene GH9/S (WP_008400439), *Escherichia coli* BL21(DE3) was grown in Luria-Bertani broth (Sigma, Dorset, UK) with 50 $\mu g.ml^{-1}$ ampicillin at 37°C, with orbital shaking at 180 rpm.

Human faecal isolates *C. eutactus* ART55/1 (Louis *et al*., 2004; note that its Genbank designation is *Coprococcus* sp. ART55/1 but for clarity it is named based on its phylogeny here as per Fig. 2), *Coprococcus* sp. L2-50 (Barcenilla *et al*., 2000; note that its Genbank designation is *Clostridium* sp. ART55/1 but for clarity it is named based on its phylogeny here as per Fig. 2), *C. eutactus* ATCC 27759, *C. comes* ATCC 27758, *C. comes* A2-232 (Barcenilla, 1999), *C. comes* SL7/1 (Louis *et al*., 2004) and *C. catus* GD/7 (Reichardt *et al*., 2014) were maintained anaerobically on M2GSC medium at 37°C (Miyazaki *et al*., 1997). Growth tests on different carbohydrates were performed in modified yeast extract-casitone-fatty acids (YCFA) medium (Duncan *et al*., 2002) in 96-well plates in an anaerobic cabinet at 37°C under 10% (v/v) carbon dioxide, 10% (v/v) hydrogen and 80% (v/v) nitrogen atmosphere with a medium pH at start of the experiment of 6.5 $\pm$ 0.2. Bacterial cellulase activity was determined using cultures grown in Hungate tubes with different cellulosic substrates. Optical density and pH readings were used as indicators for growth. For transcriptomic and proteomic analysis, overnight cultures grown in YCFA medium containing 0.2% (w/v) glucose were inoculated into YCFA containing 0.2% (w/v) of one of the following carbohydrates: glucose, cellobiose, barley β-glucan, lichenan or glucomannan and grown to exponential phase. Full details on media and growth conditions are given in supplemental methods.

Plasmid metagenomic library construction, functional screening and bioinformatic analysis

A freshly voided faecal sample was collected from a healthy female volunteer who had not received any antibiotics or other drugs during 6 months prior the sampling. The detailed method of library construction and functional screening is given in supplemental methods. Briefly, total metagenomic DNA was extracted, size fractionated and cloned into shuttle plasmid pTRKL2 (O'Sullivan and Klaenhammer, 1993). Transformation into *E. coli* XL1 Blue and selection of white colonies resulted in a library of 6146 viable clones containing an insert with an average insert size of 2.5 kb based on PCR colony screening of 24 randomly picked colonies. The library was transferred by pooling the *E. coli* XL1 Blue library and transforming the extracted DNA into *L. lactis* MG1363. It consisted of 4608 clones with 75% insert frequency and an average insert size estimated at 2.5 kb based on PCR colony screening of 15 random clones. Both libraries were arrayed on agar plates containing BHI (*E. coli*) or GM17 (*L. lactis*) medium with potato starch (Sigma-Aldrich S2004, 1% w/v), carboxymethyl cellulose (CMC, Sigma-Aldrich C4888, 0.5% w/v), lichenan (Sigma-Aldrich L6133, 0.05% w/v), oat spelt xylan (Sigma-Aldrich X0627, 0.5% w/v), polygalacturonic acid (Sigma-Aldrich P0853, 0.5% w/v), 4-methylumbelliferyl α-L-arabinofuranoside (Sigma-Aldrich M9519, 50 μg/ml) or 4-methylumbelliferyl α-L-rhamnopyranoside (Sigma-Aldrich M8412, 50 μg/ml) and incubated overnight at 37°C (*E. coli*) or 30°C (*L. lactis*). Positive clones identified by clearing zones or fluorescence were re-assessed to confirm activities and inserts sequenced (accession numbers see Supporting Information Table S1). Bioinformatic sequence analyses and databases used are detailed in supplemental methods.

Cloning of GH9 encoding genes from Coprococcus-related species

Cloning procedures are given in detail in supplemental methods. Briefly, four genes encoding putative GH9 enzymes from *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 were cloned into shuttle plasmid pTRKL2 and transformed into *E. coli* XL1 Blue and *L. lactis* MG1363. For *Coprococcus* sp. L2-50 gene GH9/S (WP_008400439), synthetic gene constructs containing different domains (Supporting Information Fig. S5) were overexpressed in *E. coli* BL21(DE3) and enzyme activities were determined as described below.

Enzyme activity assays

For GH9 constructs cloned into *E. coli* XL1 Blue and *L. lactis* MG1363, freshly grown overnight culture (10 μl) was pipetted onto the surface of lichenan-, CMC- and xylan-containing agar plates, allowed to incubate overnight and stained with Congo Red. In order to increase the contrast, the plates were flooded with 50 mM acetic acid to turn the background towards the blue colour instead of pale orange.

For determination of enzyme activities in liquid culture and cellular localisation in *E. coli* XL1 Blue and *L. lactis* MG1363, freshly grown overnight cultures of recombinant clones were analysed. The enzyme activity was determined by reducing sugar assay following the Lever method (Lever, 1977), using supernatants and cell-free extracts prepared from three independently grown cultures (for details see supplemental methods). Enzyme activity was determined by measuring the amount of reducing sugar released by the fractions incubated with CMC, lichenan or β-glucan (each at 0.5% w/v) as substrates at 37°C.

For enzyme activities in *E. coli* BL21 (DE3) containing recombinant genes, *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50, cells were harvested by centrifugation, washed in 20 ml and resuspended in 5 ml of sodium phosphate buffer (50 mM, pH 6.5). Cell extracts were prepared by bead-beating and enzyme activities determined by Lever assay (Lever, 1977). The details are provided in supplemental methods.

Gene expression and proteomic analysis of C. eutactus *ART55/1*

Full details of gene expression and proteomics methods are provided in supplemental methods. Briefly, triplicate cultures of *C. eutactus* ART55/1 grown on either glucose, cellobiose, β-glucan, lichenan or glucomannan were harvested during exponential phase (optical density 0.45–0.84, Supporting Information Table S4) and RNA and protein fractions were prepared.

For RNA sequencing, libraries were prepared after ribosomal RNA depletion and sequenced using the High Output 1X75 kit on the Illumina NextSeq 500 platform with v2 chemistry, producing 75 bp single end reads. In total, between 22 118 233 and 40 885 793 reads were produced per sample after quality filtering (99.9% of the raw read on average, Supporting Information Table S4). Raw sequencing data have been deposited in the Array Express database under the E-MTAB-8048. Reads were aligned against the reference genome for *C. eutactus* ART55/1 (FP929039.1, between 88.78% and 90.40% of the filtered reads, Supporting Information Table S4) and counted at gene locations (57.52%–60.00% counted, Supporting Information Table S4). For differential gene expression analysis, genes that had a CPM (count per million) value of more than one in three or more samples were kept for analysis, and all other genes were removed as low count genes, leaving 2022 genes for analysis. Differential expression analysis was performed using a generalized linear model with contrasts made between glucose (as the baseline) and all other carbon sources and setting significance at false discovery rate (FDR) < 0.05 and Log fold change (LogFC) > 1.

For functional analysis, significant differentially expressed gene sequences were isolated from the genome assembly and compared to the NCBI non-redundant protein database and to the InterPro protein signature database. The results of these searches were analysed with Blast2GO (version 5.2.5) (Conesa *et al*., 2005) where gene ontology (GO) terms were assigned to 1669 genes. GO enrichment analysis was carried out with Blast2GO using a Fisher Exact Test.

Protein digestion was carried out with porcine trypsin. Peptides were desalted and analysed by LC–MS as previously described (Herrero-de-Dios *et al*., 2018) using a Q Exactive Plus/Ultimate 3000RSLC nanoLC-MS system (Thermo Fisher Scientific, Hemel Hempstead, UK) to which a 25 cm long PepMap RSLC C18 nano column (internal diameter 75 μm) was fitted. Peak identification and quantification was carried out using MaxQuant (version 1.6.3.4) (Cox & Mann, 2008) with comparisons made to *C. eutactus* ART55/1 reference protein sequences downloaded from NCBI (https://www.ncbi.nlm.nih.gov/genome/13745?genome_assembly_id=175605). Parameters were set to calculate LFQ, as well as to identify potential contaminant proteins from media. In total, 919 proteins were identified by MaxQuant. Nineteen potential contaminants and 10 reverse sequence control proteins were removed before further analysis, leaving 891 proteins. LFQ values were log2 converted and then analysed for differential expression using LIMMA (version 3.38.3) (Ritchie *et al*., 2015) with an identical linear model as to that used with the RNA sequencing analysis, treating glucose as the baseline for comparisons.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al*., 2019) partner repository with the dataset identifier PXD014174.

Acknowledgements

Freda M. Farquharson, Harry J. Flint and Petra Louis received financial support from the Scottish Government Rural and Environment Science and Analytical Services Division (RESAS). Anna M. Alessi received a PhD scholarship jointly funded by the Institute of Food Research (Norwich, UK, now Quadram Institute) and the Rowett Institute (University of

Aberdeen, Aberdeen, UK). The authors wish to thank Gill Campbell and Pauline Young for their assistance with colony picking and screening metagenomic libraries, Brennan Martin at the Centre for Genome Enabled Biology and Medicine, University of Aberdeen, for performing the RNA sequencing, Evelyn Argo and Craig Pattinson for advice on sample preparation for proteomics work, Bekhal Kareem Sharif and Manon Le Merrer for help with growth experiments, Sylvia Duncan for the gift of acid-swollen cellulose and for providing strains *C. comes* A2-232 and SL7/1, Pat Bain for help with figure formats and Mike Peck and Bruce Pearson at the Institute for Food Research Norwich for helpful discussions.

References

Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* **12**: 186.

Barcenilla, A. (1999) Diversity of the butyrate-producing microflora on the human gut. PhD Thesis. Aberdeen, UK: Robert Gordon University.

Barcenilla, A., Pryde, S.E., Martin, J.C., Duncan, S.H., Stewart, C.S., Henderson, C., and Flint, H.J. (2000) Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl Environ Microbiol* **66**: 1654–1661.

Burdette, L.A., Leach, S.A., Wong, H.T., and Tullman-Ercek, D. (2018) Developing gram-negative bacteria for the secretion of heterologous proteins. *Microb Cell Fact* **17**: 196.

Cann, I., Bernardi, R.C., and Mackie, R.I. (2016) Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulose paradigm. *Environ Microbiol* **18**: 307–310.

CAZypedia Consortium. (2018) Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **28**: 3–8.

Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Dore, J., *et al*. (2013) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS ONE* **8**: e72766.

Chassard, C., Delmas, E., Robert, C., and Bernalier-Donadille, A. (2010) The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. *FEMS Microbiol Ecol* **74**: 205–213.

Chassard, C., Delmas, E., Robert, C., Lawson, P.A., and Bernalier-Donadille, A. (2011) *Ruminococcus champanellensis* sp. nov., a cellulose-degrading bacterium from the human gut microbiota. *Int J Syst Evol Microbiol* **62**: 138–143.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–1372.

Devillard, E., Goodheart, D.B., Karnati, S.K.R., Bayer, E.A., Lamed, R., Miron, J., *et al*. (2004) *Ruminococcus albus* 8 mutants defective in cellulose degradation are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. *J Bacteriol* **186**: 136–145.

Dobrijevic, D., di Liberto, G., Tanaka, K., de Wouters, T., Dervyn, R., Boudebbouze, S., *et al*. (2013) High-throughput system for the presentation of secreted and surface-exposed proteins from gram-positive bacteria in functional metagenomics studies. *PLos one* **8**: e65956.

Duncan, S.H., Barcenilla, A., Stewart, C.S., Pryde, S.E., and Flint, H.J. (2002) Acetate utilization and butyryl coenzyme a (CoA):acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl Environ Microbiol* **68**: 5186–5190.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* **6**: 121–131.

Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P., and Forano, E. (2012a) Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**: 289–306.

Flint, H.J., Scott, K.P., Louis, P., and Duncan, S.H. (2012b) The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* **9**: 577–589.

Freudl, R. (2018) Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact* **17**: 52.

Gabor, E.M., Alkema, W.B.L., and Janssen, D.B. (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* **6**: 879–886.

Gagic, D., Ciric, M., Wen, W.X., Ng, F., and Rakonjac, J. (2016) Exploring the secretomes of microbes and microbial communities using filamentous phage display. *Front Microbiol* **7**: 429.

Gasson, M.J. (1983) Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. *J Bacteriol* **154**: 1–9.

Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**: 669–685.

Herrero-de-Dios, C., Day, A., Tillmann, A.T., Kastora, S.L., Stead, D., Salgadom, P.S., *et al*. (2018) Redox regulation, rather than stress-induced phosphorylation, of a Hog1 mitogen-activated protein kinase modulates its nitrosative-stress-specific outputs. *mBio* **9**: e02229-17.

Holdeman, L.V., and Moore, W.E.C. (1974) New genus, *Coprococcus*, twelve new species, and emended descriptions of four previously described species of bacteria from human feces. *Int J Syst Bacteriol* **24**: 260–277.

Howlett, J.F., Betteridge, V.A., Champ, M., Craig, S.A.S., Meheust, A., and Jones, J.M. (2010) The definition of dietary fiber – discussions at the ninth Vahouny fiber symposium: building scientific agreement. *Food Nutr Res* **54**: 5750.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**: 1547–1549.

Kurokawa, J., Hemjinda, E., Arai, T., Kimura, T., Sakka, K., and Ohmiya, K. (2002) *Clostridium thermocellum* cellulase CelT, a family 9 endoglucanase without an Ig-like domain or family 3c carbohydrate-binding module. *Appl Microbiol Biotechnol* **59**: 455–461.

Lever, M. (1977) Carbohydrate determination with 4-hydroxybenzoic acid hydrazide (PAHBAH): effect of bismuth on the reaction. *Anal Biochem* **81**: 21–27.

Lombard, V., Golaconda Ramulu, H., Drula, E., Couthino, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucl Acids Res* **42**: D490–D495.

Louis, P., Duncan, S.H., McCrae, S.I., Millar, J., Jackson, M. S., and Flint, H.J. (2004) Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol* **186**: 2099–2106.

Louis, P., and Flint, H.J. (2017) Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol* **19**: 29–41.

McMahon, M.D., Guan, C., Handelsman, J., and Thomasa, M.G. (2012) Metagenomic analysis of *Streptomyces lividans* reveals host-dependent functional expression. *Appl Environ Microbiol* **78**: 3622–3629.

Mingardon, F., Chanal, A., Tardif, C., and Fierobe, H.P. (2011) The issue of secretion on heterologous expression of *Clostridium cellulolyticum* cellulase-encoding genes in *Clostridium acetobuylicum* ATCC 824. *Appl Environ Microbiol* **77**: 2831–2838.

Miyazaki, K., Martin, J.C., Marinsek-Logar, R., and Flint, H.J. (1997) Degradation and utilization of xylans by the rumen anaerobe *Prevotella bryantii* (formerly *P. ruminicola* subsp. *bovis*) B14. *Anaerobe* **3**: 373–381.

Moracci, M. (2019) Glycoside Hydrolase Family 42 in CAZypedia. URL http://www.cazypedia.org/.

Morais, S., Ben David, Y., Bensoussan, L., Duncan, S.H., Koropatkin, N.M., Martens, E.C., *et al*. (2016) Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, *Ruminococcus champanellensis*, reveals a fine-tuned system for cohesin-dockerin recognition. *Environ Microbiol* **18**: 542–556.

Nylund, L., Nermes, M., Isolauri, E., Salminen, S., de Vos, W.M., and Satokari, R. (2015) Severity of atopic disease inversely correlates with intestinal microbiota diversity and butyrate-producing bacteria. *Allergy* **70**: 241–244.

O'Sullivan, D.J., and Klaenhammer, T.R. (1993) High- and low-copy-number *Lactococcus* shuttle cloning vectors with features for clone screening. *Gene* **137**: 227–231.

Pereira, J.H., Sapra, R., Volponi, J.V., Kozina, C.L., Simmons, B., and Adams, P.D. (2009) Structure of endoglucanase Cel9A from the thermoacidophilic *Alicyclobacillus acidocaldarius*. *Acta Crystallogr D Biol Crystallogr* **65**: 744–750.

Pereira, J.H., Chen, Z., McAndrew, R.P., Sapra, R., Chhabra, S.R., Sale, K.L., *et al*. (2010) Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima*. *J Struct Biol* **172**: 372–379.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., *et al*. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**: D442–D450.

Pontes, D.S., De Azevedo, M.S.P., Chatel, J.-M., Langella, P., Azevedo, V., and Miyoshi, A. (2011) *Lactococcus lactis* as a live vector: heterologous protein production and DNA delivery systems. *Protein Exp Purif* **79**: 165–175.

Reichardt, N., Duncan, S.H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K.P., *et al*. (2014) Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J* **8**: 1323–1335.

Reichardt, N., Vollmer, M., Holtrop, G., Farquharson, F., Wefers, D., Bunzel, M., *et al*. (2018) Specific substrate-driven changes in human faecal microbiota composition contrast with functional redundancy in short-chain fatty acid production. *ISME J* **12**: 610–622.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl Acids Res* **43**: e47.

Robert, C., Chassard, C., Lawson, P.A., and Bernalier-Donadille, A. (2007) *Bacteroides cellulosilyticus* sp. nov., a cellulolytic bacterium from the human gut microbial community. *Int J Syst Evol Microbiol* **57**: 1516–1520.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Sheridan, P.O., Martin, J.C., Lawley, T.D., Browne, H.P., Harris, H.M.B., Bernalier-Donadille, A., *et al*. (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic firmicutes. *Microb Genom* **2**: 1–16.

Slavin, J. (2013) Fiber and prebiotics: mechanisms and health benefits. *Nutrients* **5**: 1417–1435.

Sun, Y., Su, Y., and Zhu, W. (2016) Microbiome-metabolome responses in the cecum and colon of pig to a high resistant starch diet. *Front Microbiol* **7**: 779.

Tamura, K., Nei, M., and Kumar, S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* **101**: 11030–11035.

Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., *et al*. (2010) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res* **20**: 1605–1612.

Tuncil, Y.E., Xiao, Y., Porter, N.T., Reuhs, B.L., Martens, E. C., and Hamaker, B.R. (2017) Reciprocal prioritization to dietary glycans by gut bacteria in a competitive environment promotes stable coexistence. *mBio* **8**: e01068-17.

Uchiyama, T., and Miyazaki, K. (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20**: 616–622.

Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., *et al*. (2019) The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat Microbiol* **4**: 623–632.

Vogel, C., and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**: 227–232.

Wilson, D. B., and Urbanowicz, B. (2019) Glycoside Hydrolase Family 9 in CAZypedia. URL http://www.cazypedia.org/.

Environmental Microbiology, **22,** 2150–2164

Ze, X., Ben David, Y., Laverde-Gomez, J., Dassa, B., Sheridan, P.O., Duncan, S.H., *et al*. (2015) Unique organization of extracellular amylases into 'amylosomes' in the resistant starch-utilizing human colonic Firmicutes bacterium *Ruminococcus bromii*. *mBio* **6**: e01425-15.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., *et al*. (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucl Acids Res* **46**: W95–W101.

Zhang, S., Zubay, G., and Goldman, E. (1991) Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene* **105**: 61–72.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1 Sequence analysis of clones recovered by a functional screening of human gut microbiome metagenomic libraries.

Table S2. Enzyme activity of GH9 enzymes from *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 expressed in *E. coli* XL1 Blue and *L. lactis* MG1363 supernatant and cell-free extract.

Table S3. Reducing sugar assay of cell extracts and supernatants of cultures of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 grown on glucose, cellobiose or β-glucan to early stationary phase.

Table S4: Individual culture and sample statistics of gene expression and proteomics analysis.

Table S5. Log fold change (logFC) and adjusted significance values (FDR and adj.P.Val) for RNA sequencing and proteomics analyses of *C. eutactus* ART55/1 genes.

Table S6. Significantly enriched Gene Ontology (GO) terms identified by Blast2GO (FDR < 0.05) *C. eutactus* ART55/1 genes.

Table S7. CAZyome of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50.

Table S8. Primers used for amplification of GH9 genes from *Coprococcus*-related species.

Fig. S1 Phylogenetic analysis of the family 9 catalytic modules of *C. eutactus* ART55/1 and *Coprococcus* sp. L2-50 GH9 genes (highlighted by blue and purple dots, respectively). The phylogenetic trees show the relationship of the catalytic domains of ART_GH9/S and L2-50_GH9/S proteins (A) and ART_GH9/L and L2-50_GH9/L (B) proteins. The amino acid sequences of the catalytic domains were retrieved from NCBI following BlastP analysis. The evolutionary history was inferred using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. The analysis involved 68 (A) and 44 (B) amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 376 (A) and 348 (B) positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

Fig. S2. Carbohydrate utilization of *C. eutactus* ART55/1, *Coprococcus* sp. L2-50, *C. eutactus* ATCC 27759, *C. comes* ATCC 27758, *C. comes* A2-232, *C. comes* SL7/1 and *C. catus* GD7, on different substrates.

A. Medium pH drop after growth in 96-well plates after 48 h of incubation (mean and standard deviation of triplicate cultures).

B. Growth curves during growth in 96-well plates. *C. catus* GD/7 does not grow well on glucose, but the preculture showed an optical density increase during the hour before inoculation and had reached OD 0.25 at inoculation of the growth experiment.

Fig. S3. Relationship of all differentially expressed *C. eutactus* ART55/1 genes (DEGs, 299 genes in total) between different growth substrates, expressed as logFC. Other comparisons are shown in Fig. 5C.

Fig. S4. Venn diagrams comparing significant differential expression of *C. eutactus* ART55/1 genes and proteins in glucose vs. β-glucan (A), glucose vs. cellobiose (B), glucose vs. glucomannan (C), and glucose vs. lichenan (D).

Fig. S5. Cloning strategy for generation of synthetic gene constructs of *Coprococcus* sp. L2-50 GH9/S.

Appendix S1: Supplemental methods

Received: 11 March 2021 | Revised: 12 May 2021 | Accepted: 8 June 2021

DOI: 10.1002/ijc.33739

TUMOR MARKERS AND SIGNATURES

IJC International Journal of Cancer

Genome-wide association studies of toxicity to oxaliplatin and fluoropyrimidine chemotherapy with or without cetuximab in 1800 patients with advanced colorectal cancer

Katie Watts[1] | Christopher Wills[1] | Ayman Madi[2] | Claire Palles[3] | Timothy S. Maughan[4] | Richard Kaplan[5] | Nada A. Al-Tassan[6] | Rachel Kerr[7] | David Kerr[8] | Victoria Gray[1] | Hannah West[1] | Richard S. Houlston[9] | Valentina Escott-Price[10] | Jeremy P. Cheadle[1]

[1]Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff, UK

[2]The Clatterbridge Cancer Centre NHS Foundation Trust, Bebington, UK

[3]Institute of Cancer and Genomic Sciences, Institute of Biomedical Research, University of Birmingham, Birmingham, UK

[4]CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Oxford, UK

[5]MRC Clinical Trials Unit, University College of London, London, UK

[6]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

[7]Department of Oncology, University of Oxford, Oxford, UK

[8]Nuffield Department of Clinical Laboratory Sciences, University of Oxford, John Radcliffe Hospital, Oxford, UK

[9]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

[10]Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK

Correspondence
Jeremy P. Cheadle, Division of Cancer and Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK.
Email: cheadlejp@cardiff.ac.uk

Funding information
Cancer Research UK, Grant/Award Number: C1298/A8362; Cardiff University School of

Abstract

Chemotherapies administered at normal therapeutic dosages can cause significant side-effects and may result in early treatment discontinuation. Inter-individual variation in toxicity highlights the need for biomarkers to personalise treatment. We sought to identify such biomarkers by conducting 40 genome-wide association studies, together with gene and gene set analyses, for any toxicity and 10 individual toxicities in 1800 patients with advanced colorectal cancer treated with oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab from the MRC COIN and COIN-B trials (385 patients received FOLFOX, 360 FOLFOX + cetuximab, 707 XELOX and 348 XELOX + cetuximab). Single nucleotide polymorphisms (SNPs), genes and gene sets that reached genome-wide or suggestive significance were validated in independent patient groups. We found that *MROH5* was significantly associated with neutropenia in MAGMA gene analyses in patients treated with XELOX ($P = 6.6 \times 10^{-7}$) and was independently validated in those receiving XELOX + cetuximab; pooled $P = 3.7 \times 10^{-7}$. rs13260246 at 8q21.13 was significantly associated with vomiting in patients treated with XELOX (odds ratio = 5.0, 95% confidence interval = 3.0-8.3, $P = 9.8 \times 10^{-10}$) but was not independently replicated. SNPs at 139 loci had suggestive associations for toxicities and lead SNPs at five of these were independently validated (rs6030266 with diarrhoea, rs1546161 with hand-foot syndrome, rs9601722 with neutropenia, rs13413764 with lethargy and rs4600090 with nausea; all with pooled *P*'s < 5.0×10^{-6}). In conclusion, the association of *MROH5* with neutropenia and five other putative biomarkers warrant further investigation for their potential clinical utility. Despite our comprehensive genome-wide analyses of large, well-characterised, clinical trials, we found a lack of common variants with modest effect sizes associated with toxicities.

Abbreviations: CRC, colorectal cancer; CTCAE, Common Terminology Criteria for Adverse Events; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; QUASAR2, Quick and Simple and Reliable Trial; SNP, single nucleotide polymorphism; sQTL, splicing quantitative trait loci.

Medicine, Grant/Award Number: PhD studentship; KFSHRC; Tenovus Cancer Care

KEYWORDS
chemotherapy, colorectal cancer, GWAS, toxicity

What's new?

Among cancer patients, toxic side effects of chemotherapeutic agents can vary considerably. This inter-individual variability may be influenced by genetic factors. Here, genome-wide association studies were analysed for toxicities to oxaliplatin and fluoropyrimidine chemotherapy (FOLFOX and XELOX) in patients with advanced colorectal cancer. In XELOX-treated patients, *MROH5* variants were strongly associated with neutropenia. The variant rs13260246, mapped to *SLC26A7*, was associated with vomiting. Suggestive associations for toxicities were identified for single nucleotide polymorphisms (SNPs) at 139 loci; 5 of which validated. The validated variants could serve as predictive biomarkers for specific chemotherapy-related toxicities.

1 | INTRODUCTION

Many patients diagnosed with colorectal cancer (CRC) receive chemotherapy either as part of their treatment for curative disease or to extend survival.[1] Most chemotherapeutic agents are associated with significant side effects even if administered at normal therapeutic dosages.

The combination of fluoropyrimidine and oxaliplatin is a common first-line treatment for many cancers including CRC.[2] XELOX (XEL = capecitabine, OX = oxaliplatin) is an oral fluoropyrimidine with similar efficacy to FOLFOX (FOL = folinic acid, F = fluorouracil, OX = oxaliplatin) but with differing toxicity profiles.[3,4] Whereas XELOX often causes gastrointestinal symptoms and hand-foot syndrome, FOLFOX tends to affect immunity. Cetuximab, a monoclonal antibody directed against the epidermal growth factor receptor, is also used in the treatment of CRC and often causes skin rashes.[5]

Some toxicities have short-term acute effects whereas others remain after treatment has stopped.[6] Toxicity adversely affects a patient's quality of life and can be life threatening. Drug toxicity may result in treatment discontinuation or dose reduction,[7,8] thus significantly affecting the prospects of a cure.[9,10]

Since there is significant inter-individual variation in chemotherapy-related toxicity, the identification of predictive biomarkers is highly desirable to personalise therapy. The role of inherited genetic factors is increasingly being recognised to influence patient chemotherapy-related toxicity. Notably, rare variants in the gene encoding dihydropyrimidine dehydrogenase (*DPYD*) are well established to be associated with severe toxicities to 5-fluorouracil (5-FU).[11,12] While the role of common genetic variation is less clear, we and others have shown that common variants in *DPYD* also appear to affect the toxicity.[13-15] To date, most studies have sought to identify inherited predictive biomarkers using candidate gene and variant-based analyses, based on preconceptions as to probable biology and using small cohorts of patients with no independent validation. To address such limitations, we have analysed genome-wide association study (GWAS) data on 1800 patients with advanced CRC treated with oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab with replication in independent patient groups.

2 | MATERIALS AND METHODS

2.1 | Patients and samples

In total, 2671 patients with metastatic or locally advanced colorectal adenocarcinoma were recruited into the MRC clinical trials COIN (ISRCTN27286448)[16,17] and COIN-B (ISRCTN3837568).[18] None of the patients had previously received chemotherapy for advanced disease. COIN patients were randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy (Arm A, n = 815), continuous chemotherapy with cetuximab (Arm B, n = 815) or intermittent chemotherapy (Arm C, n = 815). COIN-B patients were randomised 1:1 to receive intermittent chemotherapy and cetuximab (Arm D, n = 112) or intermittent chemotherapy and continuous cetuximab (Arm E, n = 114) (Figure 1). For the first 12 weeks, treatments were identical in all patients apart from the choice of fluoropyrimidine (n = 1068, 40% received FOLFOX and n = 1603, 60% received XELOX) together with the randomisation of ± cetuximab (n = 1041, 39% received cetuximab) (Figure 1). Overall, patients had a mean age at randomisation of 62 years (range, 18-87) and 36% were female. Blood DNA samples were prepared from 2244 of the 2671 patients.

2.2 | Clinical end points assessed and power considerations

Assessment of toxicity was performed at 12 weeks, since at this point patients from all trial arms received identical levels of chemotherapy (choice of XELOX or FOLFOX) with or without cetuximab. This time point was also prior to any interruption to treatment for the intermittent therapy arms.

The primary end point assessed was any toxicity graded by critical adverse events as per the Common Terminology Criteria for Adverse Events (CTCAE version 4.0) with the highest grade noted within the first 12 weeks of treatment (assessed at 6 and 12 weeks). Secondary end points were individual toxicities (diarrhoea, neutropenic sepsis,

Medicine, Grant/Award Number: PhD studentship; KFSHRC; Tenovus Cancer Care

KEYWORDS
chemotherapy, colorectal cancer, GWAS, toxicity

What's new?

Among cancer patients, toxic side effects of chemotherapeutic agents can vary considerably. This inter-individual variability may be influenced by genetic factors. Here, genome-wide association studies were analysed for toxicities to oxaliplatin and fluoropyrimidine chemotherapy (FOLFOX and XELOX) in patients with advanced colorectal cancer. In XELOX-treated patients, *MROH5* variants were strongly associated with neutropenia. The variant rs13260246, mapped to *SLC26A7*, was associated with vomiting. Suggestive associations for toxicities were identified for single nucleotide polymorphisms (SNPs) at 139 loci; 5 of which validated. The validated variants could serve as predictive biomarkers for specific chemotherapy-related toxicities.

1 | INTRODUCTION

Many patients diagnosed with colorectal cancer (CRC) receive chemotherapy either as part of their treatment for curative disease or to extend survival.[1] Most chemotherapeutic agents are associated with significant side effects even if administered at normal therapeutic dosages.

The combination of fluoropyrimidine and oxaliplatin is a common first-line treatment for many cancers including CRC.[2] XELOX (XEL = capecitabine, OX = oxaliplatin) is an oral fluoropyrimidine with similar efficacy to FOLFOX (FOL = folinic acid, F = fluorouracil, OX = oxaliplatin) but with differing toxicity profiles.[3,4] Whereas XELOX often causes gastrointestinal symptoms and hand-foot syndrome, FOLFOX tends to affect immunity. Cetuximab, a monoclonal antibody directed against the epidermal growth factor receptor, is also used in the treatment of CRC and often causes skin rashes.[5]

Some toxicities have short-term acute effects whereas others remain after treatment has stopped.[6] Toxicity adversely affects a patient's quality of life and can be life threatening. Drug toxicity may result in treatment discontinuation or dose reduction,[7,8] thus significantly affecting the prospects of a cure.[9,10]

Since there is significant inter-individual variation in chemotherapy-related toxicity, the identification of predictive biomarkers is highly desirable to personalise therapy. The role of inherited genetic factors is increasingly being recognised to influence patient chemotherapy-related toxicity. Notably, rare variants in the gene encoding dihydropyrimidine dehydrogenase (*DPYD*) are well established to be associated with severe toxicities to 5-fluorouracil (5-FU).[11,12] While the role of common genetic variation is less clear, we and others have shown that common variants in *DPYD* also appear to affect the toxicity.[13-15] To date, most studies have sought to identify inherited predictive biomarkers using candidate gene and variant-based analyses, based on preconceptions as to probable biology and using small cohorts of patients with no independent validation. To address such limitations, we have analysed genome-wide association study (GWAS) data on 1800 patients with advanced CRC treated with oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab with replication in independent patient groups.

2 | MATERIALS AND METHODS

2.1 | Patients and samples

In total, 2671 patients with metastatic or locally advanced colorectal adenocarcinoma were recruited into the MRC clinical trials COIN (ISRCTN27286448)[16,17] and COIN-B (ISRCTN3837568).[18] None of the patients had previously received chemotherapy for advanced disease. COIN patients were randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy (Arm A, n = 815), continuous chemotherapy with cetuximab (Arm B, n = 815) or intermittent chemotherapy (Arm C, n = 815). COIN-B patients were randomised 1:1 to receive intermittent chemotherapy and cetuximab (Arm D, n = 112) or intermittent chemotherapy and continuous cetuximab (Arm E, n = 114) (Figure 1). For the first 12 weeks, treatments were identical in all patients apart from the choice of fluoropyrimidine (n = 1068, 40% received FOLFOX and n = 1603, 60% received XELOX) together with the randomisation of ± cetuximab (n = 1041, 39% received cetuximab) (Figure 1). Overall, patients had a mean age at randomisation of 62 years (range, 18-87) and 36% were female. Blood DNA samples were prepared from 2244 of the 2671 patients.

2.2 | Clinical end points assessed and power considerations

Assessment of toxicity was performed at 12 weeks, since at this point patients from all trial arms received identical levels of chemotherapy (choice of XELOX or FOLFOX) with or without cetuximab. This time point was also prior to any interruption to treatment for the intermittent therapy arms.

The primary end point assessed was any toxicity graded by critical adverse events as per the Common Terminology Criteria for Adverse Events (CTCAE version 4.0) with the highest grade noted within the first 12 weeks of treatment (assessed at 6 and 12 weeks). Secondary end points were individual toxicities (diarrhoea, neutropenic sepsis,

self-contained models with a corrected significance threshold of $P = 5.8 \times 10^{-6}$ (Figure 1).

2.6 | Validation analyses

SNPs, genes and gene sets that reached genome-wide or suggestive significance in the GWAS analyses were independently validated in: (a) the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status and (b) the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status (Figure 1). For example, a SNP identified from the group receiving FOLFOX was validated in those receiving FOLFOX + cetuximab and in those receiving XELOX. A SNP identified from the group receiving XELOX was validated in those receiving XELOX + cetuximab and in those receiving FOLFOX. A SNP identified from the group receiving FOLFOX + cetuximab was validated in those in receiving FOLFOX and those receiving XELOX + cetuximab. A SNP identified from the group receiving XELOX + cetuximab was validated in those in receiving XELOX and those receiving FOLFOX + cetuximab (Figure 1). We considered a nominally significant threshold of $P < .05$ as evidence for validations. We had >85% power to detect our initially observed odds ratios for each validation subgroup.

Because rs13260246-reached genome-wide significance for vomiting in patients treated with XELOX, we also sought validation for this biomarker using data from 927 patients enrolled in the Quick and Simple and Reliable trial (QUASAR2). This was an open-label randomised Phase 3 clinical trial of capecitabine or capecitabine plus bevacizumab in patients with Stage II or III CRCs.[24] Patients were genotyped using the Illumina genome-wide SNP panels (Human Hap 370, Human Hap 610 or Human Omni 2.5). Imputation was performed using IMPUTEv2 with 1000 genomes as reference. The INFO score for rs13260246 was 0.96. Vomiting was graded using the CTCAE scale and patients with grades 2 to 5 (22%) were compared to those with grades 0 to 1.

2.7 | Bioinformatic analyses

The Genotype-Tissue Expression project database was used to identify expression quantitative trait loci (eQTLs) and splicing quantitative trait loci (sQTLs) for relevant SNPs (https://gtexportal.org/home). Significance for tissue association was set at $P < 1.0 \times 10^{-3}$ (ie, Bonferroni correction for 49 tissues [0.05/49]). Fine-mapping was used for SNPs at validated loci; conditional regression was first used to identify the number of causal variants and fine-mapping was then run using PAINTOR,[25] which employs a Bayesian permutation method incorporating ENCODE and FANTOM5 functional annotations. Credible sets of causal SNPs were assembled for 95% coverage.

3 | RESULTS

There were significant differences in the incidences of toxicities associated with different chemotherapy regimens and cetuximab administration in COIN and COIN-B (Table 1; Supplementary Table 1). Notably, patients treated with FOLFOX had a significantly higher incidence of neutropenic sepsis, neutropenia and stomatitis, those with

TABLE 1 Patients with grades 2 to 5 CTCAE toxicities at 12 weeks

	FOLFOX treated		XELOX treated	
	n = 385 (%)	+ cetuximab n = 360 (%)	n = 707 (%)	+ cetuximab n = 348 (%)
Any toxicity	237 (61)	275 (76)	430 (61)	226 (65)
Individual toxicities				
Diarrhoea	78 (20)	109 (30)	165 (23)	123 (35)
Neutropenic sepsis	24 (8)	39 (16)	5 (0.7)	1 (0.3)
Peripheral neuropathy	43 (11)	30 (8)	110 (16)	44 (13)
Hand-foot syndrome	9 (2)	56 (16)	53 (8)	56 (16)
Neutropenia	100 (26)	119 (33)	36 (5)	6 (2)
Lethargy	130 (34)	126 (35)	258 (36)	103 (30)
Stomatitis	48 (12)	102 (28)	32 (5)	29 (8)
Nausea	41 (11)	47 (13)	142 (20)	68 (20)
Vomiting	25 (6)	34 (9)	87 (12)	35 (10)
Rash	5 (1)	196 (54)	11 (2)	166 (48)

Note: Percentage of patients in parentheses. We had 70% power to detect a mean OR of 4.3 (range, 3-6) for any toxicity and 5.9 (2-39) for individual toxicities (Supplementary Table 3). For neutropenic sepsis in patients treated with XELOX and XELOX + cetuximab, neutropenia in patients treated with XELOX + cetuximab and rash in patients treated with FOLFOX, we had insufficient power to perform the genome-wide association studies (GWASs); therefore, in total, we conducted 40 GWASs.

XELOX had a higher incidence of nausea and those with cetuximab had a higher incidence of skin rash, hand-foot syndrome and diarrhoea (Table 1). In view of this, patients were analysed for associations with genetic biomarkers after segregation by chemotherapy treatment and cetuximab status (Figure 1). There were no clinicopathological differences between these treatment groups (Supplementary Table 2).

In total, 4 million SNPs were analysed for a relationship with any toxicity and 10 individual toxicities in each of the four patient groups. Q-Q plots of observed vs expected χ^2-test statistics showed no evidence for an inflation of test statistics for all 40 GWAS's performed (λ range, 0.99-1.02) (Supplementary Figure 1). We had 70% power to detect a mean OR of 4.3 (range, 3-6) for any toxicity and 5.9 (2-39) for individual toxicities (Supplementary Table 3).

3.1 | Relationship between SNP genotype and any toxicity

No SNPs were associated with any toxicity at genome-wide significant levels ($P < 5.0 \times 10^{-8}$). SNPs at 27 loci were associated at suggestive levels ($P < 1.0 \times 10^{-5}$) (5 with FOLFOX, 8 with FOLFOX + cetuximab, 7 with XELOX and 7 with XELOX + cetuximab) (Figure 2); however, no lead SNPs were independently validated in COIN and COIN-B patients treated with the same chemotherapy regimen but alternative cetuximab status, or alternative chemotherapy regimen but with the same cetuximab status, despite having >85% power (Supplementary Table 4).

3.2 | Relationship between SNP genotype and individual toxicity

3.2.1 | Vomiting

rs13260246 at 8q21.3 was significantly associated with vomiting in patients treated with XELOX (odds ratio [OR] = 5.0, 95% confidence intervals [CIs] = 3.0-8.3, $P = 9.8 \times 10^{-10}$; Figure 3). However, the association was not validated in COIN and COIN-B patients treated with XELOX + cetuximab ($P = .72$), nor in those receiving FOLFOX ($P = .35$), with >90% power (Supplementary Table 5). We also failed to validate the association for rs13260246 with vomiting in the QUASAR2 trial of capecitabine alone vs capecitabine + bevacizumab for Stage II and III CRC, regardless of treatment arm studied (with >99% power) (Supplementary Table 5). rs13260246 was an eQTL for *SLC26A7* and five other genes (Supplementary Figure 2). SNPs at 15 loci had suggestive associations with vomiting but none were independently validated.

3.2.2 | Diarrhoea

SNPs at 21 loci had suggestive associations with diarrhoea (Supplementary Figure 3); however, only rs6030266 at 20q13.12 in patients treated with XELOX + cetuximab (OR = 0.4, 95% CI = 0.28-0.58, $P = 5.7 \times 10^{-7}$) was validated in patients receiving FOLFOX + cetuximab (OR = 0.7, 95% CI = 0.5-0.9, $P = 3.6 \times 10^{-2}$); pooled

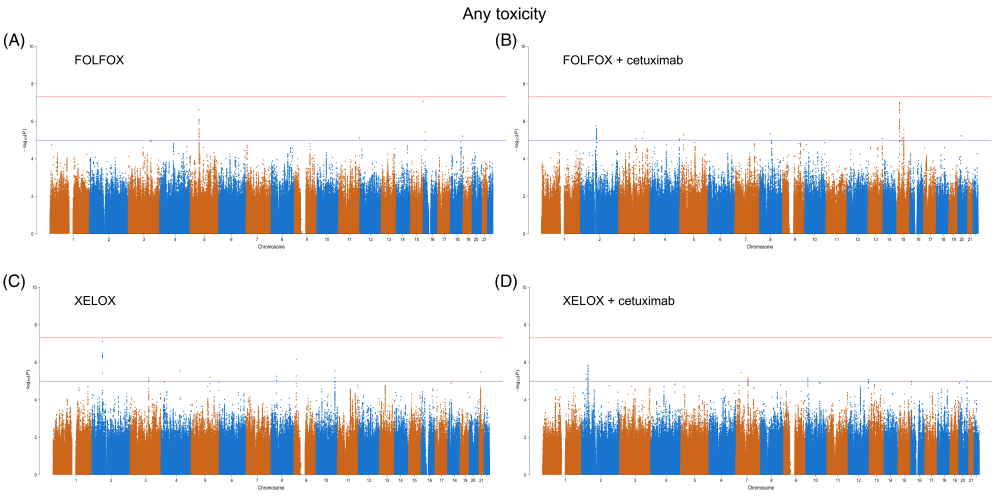

FIGURE 2 Manhattan plots of the relationship between single nucleotide polymorphism (SNP) genotype and any toxicity. Patients treated with (A) FOLFOX (n = 385), (B) FOLFOX + cetuximab (n = 360), (C) XELOX (n = 707) and (D) XELOX + cetuximab (n = 348). The red line indicates a genome-wide significance threshold of $P = 5.0 \times 10^{-8}$ and the blue line indicates a suggestive significance threshold of $P = 1.0 \times 10^{-5}$ [Color figure can be viewed at wileyonlinelibrary.com]

$P = 3.2 \times 10^{-7}$ (Table 2). rs6030266 maps to intron eight of the gene encoding protein tyrosine phosphatase receptor type T (*PTPRT*) (Supplementary Figure 4).

3.2.3 | Hand-foot syndrome

SNPs at 13 loci had suggestive associations with hand-foot syndrome (Supplementary Figure 3). Only rs1546161 at 1q21.2 in patients treated with FOLFOX (OR = 17.8, 95% CI = 5.1-62.0, $P = 5.9 \times 10^{-6}$) was validated in those receiving XELOX (OR = 1.7, 95% CI = 1.1-2.7, $P = 2.5 \times 10^{-2}$); pooled $P = 2.5 \times 10^{-6}$ (Table 2). rs1546161 maps to B-cell lymphoma 9 (*BCL9*) and was an eQTL for *GJA5* (Supplementary Figure 4).

3.2.4 | Neutropenia

SNPs at 13 loci had suggestive associations with neutropenia (Supplementary Figure 3). Only rs9601722 at 13q31.1 in patients treated with FOLFOX + cetuximab (OR = 3.4, 95% CI = 2.0-5.7, $P = 5.2 \times 10^{-6}$) was independently validated in those receiving FOLFOX (OR = 1.7, 95% CI = 1.1-2.9, $P = 3.6 \times 10^{-2}$); pooled $P = 3.0 \times 10^{-6}$ (Table 2). rs9601722 maps to a lncRNA (*LOC105370284*).

3.2.5 | Lethargy

SNPs at 12 loci had suggestive associations with lethargy (Supplementary Figure 3); however, only rs13413764 at 2q14.3 in

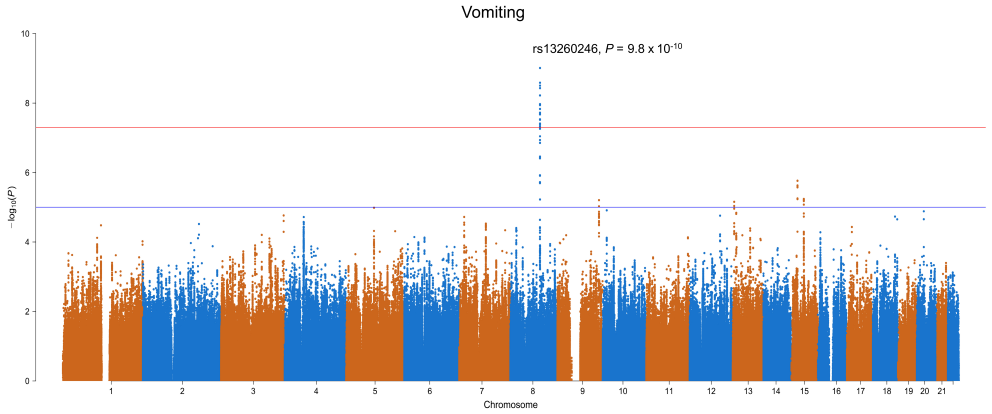

FIGURE 3 Manhattan plot of the association between single nucleotide polymorphism (SNP) genotype and vomiting in patients treated with XELOX. The red line corresponds to a $P = 5.0 \times 10^{-8}$ and the blue line $P = 1.0 \times 10^{-5}$ [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Validated single nucleotide polymorphisms (SNPs) associated with individual toxicities

				Initial GWAS			Validation chemo	Validation cetuximab status	Combined
Toxicity	**Treatment group**	**Lead SNP**	**Cytoband**	**OR**	**95% CI**	***P*-value**	***P*-value**	***P*-value**	***P*-value**
Diarrhoea	XELOX + cetuximab	rs6030266	20q13.12	0.4	0.3-0.6	5.7×10^{-7}	.33	3.6×10^{-2}	3.2×10^{-7}
Hand-foot syndrome	FOLFOX	rs1546161	1q21.2	17.8	5.1-62	5.9×10^{-6}	.13	2.5×10^{-2}	2.5×10^{-6}
Neutropenia	FOLFOX + cetuximab	rs9601722	13q31.1	3.4	2.0-5.7	5.2×10^{-6}	3.6×10^{-2}	NA	3.0×10^{-6}
Lethargy	XELOX	rs13413764	2q14.3	1.8	1.4-2.3	4.5×10^{-6}	NA	9.2×10^{-3}	7.5×10^{-7}
Nausea	FOLFOX + cetuximab	rs4600090	1p33	4.0	2.2-7.2	5.9×10^{-6}	4.2×10^{-2}	.55	4.0×10^{-6}

Abbreviations: CI, confidence intervals; Combined, pooled *P*-value of initial GWAS cohort and validated cohort (excludes cohort which was not validated); NA, OR in the opposite direction to the initial GWAS; OR, odds ratio; Validation cetuximab status, validation in the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status; Validation chemo, validation in the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status.

patients treated with XELOX (OR = 1.8, 95% CI = 1.4-2.3, $P = 4.5 \times 10^{-6}$) was replicated in those receiving FOLFOX (OR = 1.5, 95% CI = 1.1-2.1, $P = 9.2 \times 10^{-3}$); pooled $P = 7.5 \times 10^{-7}$ (Table 2). rs13413764 maps to an intergenic region.

3.2.6 | Nausea

SNPs at 12 loci had suggestive associations with nausea (Supplementary Figure 3). However, only rs4600090 at 1p33 in patients treated with FOLFOX + cetuximab (OR = 4.0, 95% CI = 2.2-7.2, $P = 5.9 \times 10^{-6}$) was independently validated in those receiving FOLFOX (OR = 2.0, 95% CI = 1.1-4.0, $P = 4.2 \times 10^{-2}$); pooled $P = 4.0 \times 10^{-6}$ (Table 2). rs4600090 was an eQTL for *CMPK1*, *FOXE3* and *PDZK1IP1* (Supplementary Figure 4).

3.2.7 | Peripheral neuropathy, stomatitis, rash and neutropenic sepsis

SNPs at 15, 10, 8 and 4 loci had suggestive associations with peripheral neuropathy, stomatitis, skin rash and neutropenic sepsis, respectively, but no lead SNPs were independently validated.

3.3 | MAGMA gene and pathway analyses

Gene and pathway analyses were performed considering approximately 17 000 genes and 8500 gene sets. Four genes were significantly associated with neutropenia (using a Bonferroni corrected threshold of $P < 2.5 \times 10^{-6}$). Of these, Maestro Heat-Like Repeat Family Member 5 (*MROH5*), found in patients treated with XELOX ($P = 6.6 \times 10^{-7}$), was independently validated in those receiving XELOX + cetuximab ($P = 3.3 \times 10^{-2}$); pooled $P = 3.7 \times 10^{-7}$ (Table 3; Supplementary Figure 5). Under a multivariate model accounting for sex and age, *MROH5* remained significant in a pooled analysis of patients treated with XELOX and XELOX + cetuximab; pooled $P = 1.0 \times 10^{-6}$.

MROH5 lies at 8q24.3, one of the 13 loci of suggestive association with neutropenia. The association of *MROH5* with neutropenia appeared to be due to independent sets of SNPs in patients treated with XELOX (lead SNP rs76380775 OR = 4.8, 95% CI = 2.4-9.5, $P = 1.4 \times 10^{-6}$) as compared to those receiving XELOX + cetuximab (lead SNP rs12056882 OR = 4.4, 95% CI = 1.4-14, $P = 1.0 \times 10^{-2}$; Supplementary Figure 6). Neither rs76380775 nor rs12056882 was associated with neutropenic sepsis or white blood cell count. rs12056882 was a sQTL for *PTP4A3* (which lies 1.37 kb downstream of *MROH5*).

One gene was significantly associated with stomatitis, 3 genes (all mapping to 8q21.3) were associated with vomiting (Table 3) and 4, 8 and 3 gene sets were associated with any toxicity, lethargy and vomiting, respectively; however, all failed independent validation (Supplementary Tables 6 and 7).

3.4 | Lack of confounding effect for rare *DPYD* variants

We have previously shown that two rare variants in *DPYD* (Asp949Val and IVS14+1G>A) were associated with a range of toxicities in COIN and COIN-B.[15] Of the 1800 patients in our current GWASs, 22 carried Asp949Val and 17 carried IVS14 +1G>A. Excluding these patients made no significant differences to the strengths of associations reported *herein* (Supplementary Table 8).

TABLE 3 MAGMA gene analyses for individual toxicities

Toxicity	Treatment group	Gene	*P*-value	Validation chemo *P*-value	Validation cetuximab status *P*-value	Pooled *P*-value
Neutropenia	FOLFOX	*RPL17-C18orf32*	8.9×10^{-7}	.57	.53	–
		C18orf32	1.3×10^{-6}	.56	.51	–
		RPL17	1.5×10^{-6}	.56	.52	–
	XELOX	***MROH5***	$\mathbf{6.6 \times 10^{-7}}$	$\mathbf{3.3 \times 10^{-2}}$	**.09**	$\mathbf{3.7 \times 10^{-7}}$
Stomatitis	FOLFOX	*SCAF4*	1.3×10^{-6}	.07	.61	–
Vomiting	XELOX	*LRRC69*	1.2×10^{-7}	.77	.73	–
		SLC26A7	4.3×10^{-7}	.81	.60	–
		PIP4P2	9.7×10^{-7}	.94	.34	–

Note: Significance was set at a Bonferroni-corrected significance threshold of $P < 2.5 \times 10^{-6}$. Only *MROH5* was significantly associated with neutropenia in patients treated with XELOX and was independently validated in patients receiving XELOX + cetuximab ($P = 3.3 \times 10^{-2}$), with a pooled $P = 3.7 \times 10^{-7}$ (in bold) (and $P = 5.8 \times 10^{-7}$ when also including the FOLFOX cohort).
Abbreviations: Validation cetuximab status, Validation in the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status; Validation chemo, validation in the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status.

3.5 | Alternative model of toxicity

We considered an alternative model of toxicity comparing patients with grades 3 to 5 (ie, severe toxicity) to patients with grades 0 to 2 (no, mild or moderate toxicity) for all biomarkers identified *herein* (Supplementary Table 9). Five of the seven biomarkers remained nominally significant.

3.6 | Evaluation of previously purported associations

A previous GWAS for toxicity to 5-FU or FOLFOX in patients with CRC identified two SNPs associated with mucositis, two with diarrhoea and three with haematological toxicities, albeit only at nominal significance.[26] We failed to validate any of these SNPs in COIN and COIN-B (Supplementary Table 10), despite having adequate power.

4 | DISCUSSION

MROH5 was identified from MAGMA gene analyses as associated with neutropenia at genome-wide significant levels in patients treated with XELOX and was independently validated in those receiving XELOX + cetuximab. Interestingly, this association appeared to be due to independent sets of SNPs in these two patient groups and rs12056882 was a sQTL for *PTP4A3* which lies adjacent to *MROH5*. *MROH5* has been suggested to be both a pseudogene and a functional gene (with an unknown role) dependent upon the status of a SNP that introduces a premature termination codon. *PTP4A3* represents a strong causal candidate for neutropenia as treatment of mice with a *PTP4A3* derived peptide reduced endotoxemia-induced septic shock.[27] *PTP4A3* expression has also been associated with poor prognosis in CRC possibly due to a role in metastasis and tumour invasion,[28,29] and has been implicated in resistance to chemotherapy.[30,31] Importantly, the strength of the relationship between SNPs in *MROH5* and neutropenia suggests that they may have clinical utility as predictive biomarkers.

We also found a clear signal for rs13260246 associated with vomiting in patients treated with XELOX. However, this association was not validated in patients treated with XELOX + cetuximab, nor in those receiving FOLFOX, nor in patients treated with capecitabine ± bevacizumab from the QUASAR2 trial. Given that we had sufficient power to replicate the initial observation, these data suggest that rs13260246 is a false-positive although it remains possible that the association with vomiting is specific to those treated with XELOX alone. rs13260246 maps to, and is an eQTL for, *SLC26A7*, which functions as a Cl^-/HCO_3^- exchanger and chloride channel,[32] and is expressed in several tissues including the thyroid. Chemotherapy can cause thyroid dysfunction and response to treatment may be affected by pre-existing thyroid conditions.[33-35] *SLC26A7* is also expressed in parietal cells and genetic deletion results in decreased gastric acid secretion.[36,37] Both thyroid and gastric dysfunction can cause vomiting.[38,39] Therefore, *SLC26A7* represents a strong biological candidate for vomiting, but lacks genetic validation.

In total, we found SNPs at 139 loci with evidence for suggestive associations for any toxicity or individual toxicities and lead SNPs at five of these were validated at nominally significant levels. However, if we applied a more stringent correction for 139 validation tests, none of the five would have passed the adjusted significance threshold. Further validation of these biomarkers in independent cohorts is therefore necessary before they could be applied in clinical practice. rs6030266 was associated with diarrhoea and identified in patients treated with cetuximab. It maps to intron eight of *PTPRT*, a tumour suppressor gene that functions as part of the JAK/STAT pathway.[40] rs1546161 was associated with hand-foot syndrome and maps to *BCL9*, overexpression of which has been linked to disrupted *wnt* signalling.[41] rs1546161 is also an eQTL for *GJA5*, a gap junction protein with significant expression in subcutaneous adipose tissue. rs4600090 associated with nausea lies within and is an eQTL for *CMPK1*, an enzyme associated with activation of 5-FU phosphorylation and linked to 5-FU sensitivity.[42] rs4600090 is also an eQTL for *PDZK1IP1* which functions as a cargo protein expressed in the adrenal glands. Interestingly, noradrenaline and cortisol, hormones released by adrenal glands, have both been associated with chemotherapy-induced nausea.[43] rs9601722 associated with neutropenia and rs13413764 with lethargy did not lie within protein coding gene regions.

Our study had limited power to detect common variants associated with toxicity with low odds ratios (<2) and our attempts to validate any findings were limited by groups with similar, but non-identical, therapies. Nonetheless, after conducting 40 GWASs on large patient cohorts with well-characterised clinical data, we conclude there is a lack of common variants with modest or large effect sizes associated with toxicities induced by oxaliplatin and fluoropyrimidine chemotherapy with or without cetuximab. In support of this, we failed to replicate loci previously suggested to be associated with toxicity to FOLFOX identified from another GWAS.[26] Further analyses of *MROH5* and/or *PTP4A3* with neutropenia are warranted.

ACKNOWLEDGEMENTS

We thank Hywel Williams for helpful advice and the patients and their families who participated and gave their consent for this research, and the investigators throughout the UK who submitted samples for assessment. This work was supported by Cardiff University School of Medicine (to K.W.), Tenovus Cancer Care (to J.P.C. and C.W.), KFSHRC (to N.A.A.) and Cancer Research UK (grant number C1298/A8362 to R.S.H.). The COIN and COIN-B trials were funded by Cancer Research UK, the Medical Research Council and an unrestricted educational grant from Merck-Serono and were conducted with the support of the National Institute of Health Research Cancer Research Network. None of the sponsors played a role in the study design; the collection, analysis and interpretation of data; the writing of the report; and the decision to submit the paper for publication.

CONFLICT OF INTEREST

T.S.M. received research funding from Merck KgAa (manufacturer of cetuximab) for unrelated research. The original COIN trial received

funding from Merck KgAa as well as Cancer Research UK. T.S.M. received honorarium and travel support from Merck KgAa. D.K. is a director of Oxford Cancer Biomarkers. All other authors have declared no potential conflicts of interest.

ETHICS STATEMENT

Patients were recruited from the MRC clinical trials COIN (ISRCTN27286448) and COIN-B (ISRCTN3837568) and all gave fully informed consent for bowel cancer research (approved by REC [04/MRE06/60]).

DATA AVAILABILITY STATEMENT

The GWAS summary statistics are available through the NHGRI-EBI GWAS Catalog under study accession numbers GCST90017191 - GCST90017231: http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90017001-GCST90018000. Further details and other data that support the findings of this study are available from the corresponding author upon request.

ORCID

Jeremy P. Cheadle https://orcid.org/0000-0001-9453-8458

REFERENCES

1. Schmoll HJ, Van Cutsem E, Stein A, et al. Esmo consensus guidelines for management of patients with colon and rectal cancer. A personalized approach to clinical decision making. *Ann Oncol*. 2012;23:2479-2516.
2. Stein A, Arnold D. Oxaliplatin: a review of approved uses. *Expert Opin Pharmacother*. 2012;13:125-137.
3. Ducreux M, Bennouna J, Hebbar M, et al. Efficacy and safety findings from a randomized phase III study of capecitabine (X) + oxaliplatin (O) (XELOX) vs. infusional 5- FU/LV + O (FOLFOX-6) for metastatic colorectal cancer (MCRC). *J Clin Oncol*. 2007;25:4029-4029.
4. Guo Y, Xiong BH, Zhang T, Cheng Y, Ma L. XELOX vs. FOLFOX in metastatic colorectal cancer: an updated meta-analysis. *Cancer Invest*. 2016;34:94-104.
5. Petrelli F, Borgonovo K, Barni S. The predictive role of skin rash with cetuximab and panitumumab in colorectal cancer patients: a systematic review and meta-analysis of published trials. *Target Oncol*. 2013;8:173-181.
6. Andreyev HJN, Davidson SE, Gillespie C, Allum WH, Swarbrick E. Practice guidance on the management of acute and chronic gastrointestinal problems arising as a result of treatment for cancer. *Gut*. 2012;61:179-192.
7. Blumenthal GM, Gong Y, Kehl K, et al. Analysis of time-to-treatment discontinuation of targeted therapy, immunotherapy, and chemotherapy in clinical trials of patients with non-small-cell lung cancer. *Ann Oncol*. 2019;30:830-838.
8. Koedoot CG, De Haan RJ, Stiggelbout AM, et al. Palliative chemotherapy or best supportive care? A prospective study explaining patients' treatment preference and choice. *Br J Cancer*. 2003;89:2219-2226.
9. Huitema ADR, Spaander M, Mathôt RAA, et al. Relationship between exposure and toxicity in high-dose chemotherapy with cyclophosphamide, thiotepa and carboplatin. *Ann Oncol*. 2002;13:374-384.
10. Braun MS, Seymour MT. Balancing the efficacy and toxicity of chemotherapy in colorectal cancer. *Ther Adv Med Oncol*. 2011;3:43-52.
11. Schwab M, Zanger UM, Marx C, et al. Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU toxicity study group. *J Clin Oncol*. 2008;26:2131-2138.
12. Henricks LM, Lunenburg CATC, de Man FM, et al. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *Lancet Oncol*. 2018;19:1459-1467.
13. Gonzalez FJ, Fernandez-Salguaro P. Diagnostic analysis, clinical importance and molecular basis of dihydropyrimidine dehydrogenase deficiency. *Trends Pharmacol Sci*. 1995;16:325-327.
14. Wei X, McLeod HL, McMurrough J, Gonzalez FJ, Fernandez-Salguero P. Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. *J Clin Invest*. 1996;98:610-615.
15. Madi A, Fisher D, Maughan TS, et al. Pharmacogenetic analyses of 2183 patients with advanced colorectal cancer; potential role for common dihydropyrimidine dehydrogenase variants in toxicity to chemotherapy. *Eur J Cancer*. 2018;102:31-39.
16. 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;377:2103-2114.
17. 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;12:642-653.
18. 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;15:631-639.
19. Moore CM, Jacobson SA, Fingerlin TE. Power and sample size calculations for genetic association studies in the presence of genetic model misspecification. *Hum Hered*. 2019;84:256-271. https://doi.org/10.1159/000508558
20. 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.
21. 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;81:559-575.
22. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw*. 2018;3:731.
23. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11:e1004219.
24. Kerr RS, Love S, Segelov E, et al. Adjuvant capecitabine plus bevacizumab versus capecitabine alone in patients with colorectal cancer (QUASAR 2): an open-label, randomised phase 3 trial. *Lancet Oncol*. 2016;17:1543-1557.
25. Kichaev G, Yang W-Y, Lindstrom S, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet*. 2014;10:e1004722.
26. Fernandez-Rozadilla C, Cazier JB, Moreno V, et al. Pharmacogenomics in colorectal cancer: a genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. *Pharmacogenomics J*. 2013;13:209-217.
27. Tang X, Woodward T, Amar S. A PTP4A3 peptide PIMAP39 modulates TNF-alpha levels and endotoxic shock. *J Innate Immun*. 2009;2:43-55.
28. Zimmerman MW, Homanics GE, Lazo JS. Targeted deletion of the metastasis-associated phosphatase Ptp4a3 (PRL-3) suppresses murine colon cancer. *PLoS One*. 2013;8:e58300.
29. Saha S, Bardelli A, Buckhaults P, et al. A phosphatase associated with metastasis of colorectal cancer. *Science*. 2001;294:1343-1346.
30. Csoboz B, Gombos I, Tatrai E, et al. Chemotherapy induced PRL3 expression promotes cancer growth via plasma membrane remodeling and specific alterations of caveolae-associated signaling. *Cell Commun Signal*. 2018;16:51.

31. den Hollander P, Rawls K, Tsimelzon A, et al. Phosphatase PTP4A3 promotes triple-negative breast cancer growth and predicts poor patient survival. *Cancer Res*. 2016;76:1942-1953.
32. Kim KH, Shcheynikov N, Wang Y, Muallem S. SLC26A7 is a Cl- channel regulated by intracellular pH. *J Biol Chem*. 2005;280:6463-6470.
33. Fujiwara Y, Chayahara N, Mukohara T, et al. Hypothyroidism in patients with colorectal carcinoma treated with fluoropyrimidines. *Oncol Rep*. 2013;30:1802-1806.
34. Andreyev HJN, Lalji A, Mohammed K, et al. The FOCCUS study: a prospective evaluation of the frequency, severity and treatable causes of gastrointestinal symptoms during and after chemotherapy. *Support Care Cancer*. 2021;29(3):1443-1453.
35. Hartmann PA-CK. Thyroid disorders in the oncology patient. *J Adv Pract Oncol*. 2015;*6*:99.
36. Xu J, Song P, Nakamura S, et al. Deletion of the chloride transporter Slc26a7 causes distal renal tubular acidosis and impairs gastric acid secretion. *J Biol Chem*. 2009;284:29470-29479.
37. Petrovic S, Ju X, Barone S, et al. Identification of a basolateral $Cl^-/HCO\ 3^-$ exchanger specific to gastric parietal cells. *Am J Physiol Liver Physiol*. 2003;284:G1093-G1103.
38. Sweet C, Sharma A, Lipscomb G. Recurrent nausea, vomiting and abdominal pain due to hypothyroidism. *BMJ Case Rep*. 2010;2010:bcr1120092461.
39. Raufman JP, Collins SM, Pandol SJ, et al. Reliability of symptoms in assessing control of gastric acid secretion in patients with Zollinger-Ellison syndrome. *Gastroenterology*. 1983;84:108-113.
40. Hsu HC, Lapke N, Chen SJ, et al. PTPRT and PTPRD deleterious mutations and deletion predict bevacizumab resistance in metastatic colorectal cancer patients. *Cancers (Basel)*. 2018;10:314.
41. Takada K, Zhu D, Bird GH, et al. Targeted disruption of the BCL9/β-catenin complex inhibits oncogenic Wnt signaling. *Sci Transl Med*. 2012;4:148ra117.
42. Yasuno H, Kurasawa M, Yanagisawa M, Sato Y, Harada N, Mori K. Predictive markers of capecitabine sensitivity identified from the expression profile of pyrimidine nucleoside-metabolizing enzymes. *Oncol Rep*. 2013;29:451-458.
43. Fredrikson M, Hursti TJ, Steineck G, Fürst CJ, Börjesson S, Peterson C. Delayed chemotherapy-induced nausea is augmented by high levels of endogenous noradrenaline. *Br J Cancer*. 1994;70:642-645.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Watts K, Wills C, Madi A, et al. Genome-wide association studies of toxicity to oxaliplatin and fluoropyrimidine chemotherapy with or without cetuximab in 1800 patients with advanced colorectal cancer. *Int. J. Cancer*. 2021;149(9):1713-1722. https://doi.org/10.1002/ijc.33739