Genome-wide analyses to identify biomarkers of toxicity in the treatment of advanced colorectal cancer

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

Katie Watts

2023

Division of Cancer and Genetics
School of Medicine
Cardiff University

Jeremy Cheadle
Valentina Escott-Price
Hywel Williams

Summary

Background

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. Inherited genetic variants are increasingly being recognised to cause chemotherapy-induced toxicity.

Aim

I sought such biomarkers by conducting genome-wide association studies, together with gene and gene set analyses, for ten toxicities in 1800 patients with advanced colorectal cancer (CRC) treated with oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab.

Materials and Methods

Patients were from the MRC COIN and COIN-B trials. 385 received folinic acid, fluorouracil and oxaliplatin (FOLFOX), 360 FOLFOX + cetuximab, 707 capecitabine and oxaliplatin (XELOX) and 348 XELOX + cetuximab. Common and low-frequency single nucleotide polymorphisms (SNPs), genes and gene sets that reached genome-wide or suggestive significance were replicated in independent patient groups, clinical trial cohorts and participants from the UK Biobank and Genomics England. Meta-analyses were also performed to increase power.

Results

rs13260246 at 8q21.13 was significantly associated with vomiting in patients treated with XELOX (Odds Ratio [OR]=5.0, 95% Confidence Interval [CI]=3.0-8.3, P=9.8x10^{-10}) but failed independent replication. SNPs at 139 loci had suggestive associations for toxicities and lead SNPs at five were replicated. rs6783836 in *ST6GAL1* was associated with hand-foot syndrome (HFS) in patients treated with XELOX (OR=3.1, 95% CI=2.1-4.6, P=4.3x10^{-8}) and *ST6GAL1* was associated with type-2 diabetes (a risk factor for HFS). A low-frequency nonsynonymous variant in the antigen processing 1 signature region was suggestive of an association with sepsis (OR=6.1, 95% CI=3.0-12.8, P=1.2x10^{-6}). rs4760830 in *TRHDE* was associated with diarrhoea in patients treated with capecitabine (OR=0.6, 95% CI=0.50-0.72, P=4.8x10^{-8}). In MAGMA gene analyses, *MROH5* was significantly associated with neutropenia (P=6.6x10^{-7}) and was independently replicated.

Conclusion

My comprehensive study has identified several biomarkers that warrant further investigation for their potential clinical utility.

Table of contents

List of Figures

List of tables

Abbreviations used throughout this thesis

ABBREVIATION	DESCRIPTION
%	Percent
5FU	Fluorouracil
95% CI	95% Confidence intervals
ARCCA	Advanced Research Computing at Cardiff
BCL9	BCL9 Transcription Coactivator
BRAF	v-Raf murine sarcoma viral oncogene homolog B
C.	Coding
CADD	Combined Annotation Dependent Depletion
Chr	Chromosome
CI	Confidence interval
CMPK1	Cytidine/Uridine Monophosphate Kinase 1
COIN	COntinuous versus INtermittent
CRAN	Comprehensive R Archive Network
CRC	Colorectal cancer
CTCAE	Common terminology criteria for adverse events
DPD / *DPYD*	Dihydropyrimidine dehydrogenase (Enzyme / Gene)
EGFR	Epidermal growth factor receptor
EMA	European Medicines Agency
ENOSF1	Enolase superfamily member 1
eQTL	Expression quantitative trait loci
FDA	US Food and Drug Administration
FOLFOX	Folinic acid, fluorouracil, oxaliplatin (chemotherapy regimen)
GEL	Genomics England
GJA5	Gap Junction Protein Alpha 5
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study/studies
HBA1C	Glycated haemoglobin
HFS	Hand foot syndrome
HLA	Human leukocyte antigen
HPC	High-performance computer
HR	Hazard ratio
HWE	Hardy-Weinberg Equilibrium
ICD	International classification of diseases
Indel	Insertion and deletion variant
KRAS	Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LD	Linkage disequilibrium
MAF	Minor allele frequency
ML	Machine learning
mRNA	Messenger RNA
MROH5	Maestro heat like repeat family member 5
MTHFR	Methylenetetrahydrofolate reductase
N	Number
NA	Not applicable
NCBI	National Center for Biotechnology Information
NHS	National Health Service

NICE	National Institute for Health and Care Excellence
NRAS	*Neuroblastoma RAS viral oncogene homolog*
OPCS4	Classification of operations and procedures
OR	Odds ratio
OS	Overall survival
P	*P*-value
P.	Protein
PCA	Principle Component Analyses
PDZK1IP1	PDZK1 Interacting Protein 1
PFS	Progression free survival
POLYPHEN2	Polymorphism Phenotyping version 2
PTP4A3	Protein tyrosine phosphatase 4A3
PTPRT	Protein tyrosine phosphatase receptor type T
QTL	Molecular quantitative trait loci
QUASAR2	Quick and Simple and Reliable Trial
QQ	Quantile-quantile
QC	Quality control
RAS	*KRAS* and *NRAS*
RECIST	Response Evaluation Criteria In Solid Tumours
SD	Standard deviation
SIFT	Sorting Intolerant from Tolerant
SNP	Single nucleotide polymorphism
STREGA	Strengthening the Reporting of Genetic Association Studies
ST6GAL1	ST6 beta-galactoside alpha-2,6-sialyltransferase 1
SQTL	Splicing quantitative trait loci
T2D	Type 2 diabetes
TAP1	Transporter associated with antigen processing 1
TPM	Transcripts per million
TRHDE	TRH degrading ectoenzyme (gene)
TRH-DE	TRH degrading ectoenzyme (enzyme)
TYMS	Thymidylate Synthetase
T2D	Type 2 diabetes
UGT1A1	UDP glucuronosyltransferase family 1 member A1
UK	United Kingdom
USA	United States of America
VEGF	Vascular endothelial growth factor
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organisation
XELOX / CAPOX	Oxaliplatin and capecitabine (chemotherapy regime)

Publications

Publications as a direct result of this work:

Watts K, Wills C, Madi A, Palles C, Maughan TS, Kaplan R, Al-Tassan NA, Kerr R, Kerr D, Gray V, West H, Houlston RS, Escott-Price V, Cheadle JP. 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 Nov 1;149(9):1713-1722. doi: 10.1002/ijc.33739.

Watts K, Wills C, Madi A, Palles C, Maughan TS, Kaplan R, Al-Tassan NA, Kerr R, Kerr DJ, Houlston RS, Escott-Price V, Cheadle JP. Genetic variation in ST6GAL1 is a determinant of capecitabine and oxaliplatin induced hand-foot syndrome. Int J Cancer. 2022 Sep 15;151(6):957-966. doi: 10.1002/ijc.34046.

Publications due to additional work I have been involved with during my PhD:

Wills C, He Y, Summers MG, Lin Y, Phipps AI, **Watts K**, Law PJ, Al-Tassan NA, Maughan TS, Kaplan R, Houlston RS, Peters U, Newcomb PA, Chan AT, Buchanan DD, Gallinger S, Marchand LL, Pai RK, Shi Q, Alberts SR, Gray V, West HD, Escott-Price V, Dunlop MG, Cheadle JP. A genome-wide search for determinants of survival in 1926 patients with advanced colorectal cancer with follow-up in over 22,000 patients. Eur J Cancer. 2021 Dec;159:247-258. doi: 10.1016/j.ejca.2021.09.047.

Wills C, **Watts K,** Maughan TS, Fisher D, Al-Tassan NA, Houlston RS, Escott-Price V, Cheadle JP. Germline variation in RASAL2 may predict survival in patients with RAS-activated colorectal cancer. Genes Chromosomes Cancer. 2023 Feb 15. doi: 10.1002/gcc.23133. Epub ahead of print.

Acknowledgements

I would like to thank the following:

Cardiff University School of Medicine for funding my PhD Scholarship. All the participants of COIN and COIN-B, without whom this project would not have been possible.

My supervisor Jeremy Cheadle for his patience, feedback, and support. My co-supervisors Valentina Escott-Price and Hywel Williams for their guidance throughout.

Chris Wills, Amy Houseman, Dr Victoria Gray, Dr Matt Summers, and Dr Hannah West for their moral support throughout my PhD journey and for keeping me entertained both in the office and virtually.

My sister Louise and brother-in-law Fred for both keeping me in spirits and for believing in me, even when I had doubts.

Finally, the biggest thanks to my parents for their unwavering support and love. Their faith in me has enabled me to heights I never thought I would achieve.

1 Introduction

1.1 Colorectal cancer

Colorectal cancer (CRC) is the third most common cancer worldwide and the second leading cause of cancer death (Baidoun *et al*, 2021). At diagnosis, approximately 25% of patients present with metastatic disease and a further 25% of patients will develop metastatic disease (Biller and Schrag, 2021). Although newer drugs have doubled CRC survival rates over the past two decades, prognosis is still poor. Only around 35% of patients with metastatic disease survive for 3 years after diagnosis, and less than 20% survive for more than 5 years (Brody, 2015; Biller and Schrag, 2021).

The majority of CRC cases develop through the adenoma-carcinoma sequence over the course of 10-15 years (Binefa *et al*, 2014). CRC is usually only symptomatic at advanced stages, which is why mortality rates are high (Dekker *et al*, 2019). Risk factors for CRC include a low-fruit high-fat diet, obesity, a sedentary lifestyle, excess alcohol intake, being male, smoking and age (Martinez *et al*, 2007; Wolin *et al*, 2009; Giovannucci, 2004; Liang *et al*, 2009). Historically, CRC has been limited to older age adults, however, in recent years there has been a rise in the number of 20-30 year olds presenting with CRC (Vuik *et al*, 2019). It is predicted that >65% of CRC cases could be avoided with a healthy lifestyle (Giovannucci, 2002).

However, CRC also has a strong hereditary component. Around 5-10% of CRC cases are due to inherited syndromes (Macaron *et al*, 2015). Most syndromes are due to mutations in either tumour suppressor or repair and stability genes (Toma *et al*, 2012). For patients with known mutations or at high risk due to a family history of

CRC, enhanced monitoring and prophylaxis can be implemented to prevent CRC development (Monahan *et al*, 2020). However, regardless of whether CRC is inherited or sporadic, treatment options remain the same (**Figure 1.1**).

1.2 CRC treatments

For the 75% of patients with localised disease at diagnosis, first line treatment is surgical resection with intent to cure (Gustavsson *et al*, 2015). For the 25% of patients with metastatic disease at diagnosis and the 45% who experience recurrence after surgical resection, a combination of chemotherapies and targeted therapies is administered (Kuipers et al, 2015). Which therapies are used depends on clinical and genetic factors, notably *RAS* and *BRAF* mutation status (Modest *et al*, 2019).

1.2.1 Chemotherapies

Chemotherapy is the most common form of anti-cancer treatment and functions by inducing cell death in a nonspecific manner (Kummar *et al*, 2006). It is administered for both curative and palliative care (Neugut and Prigerson, 2017). In the UK, several treatments are available depending on patient characteristics and choice.

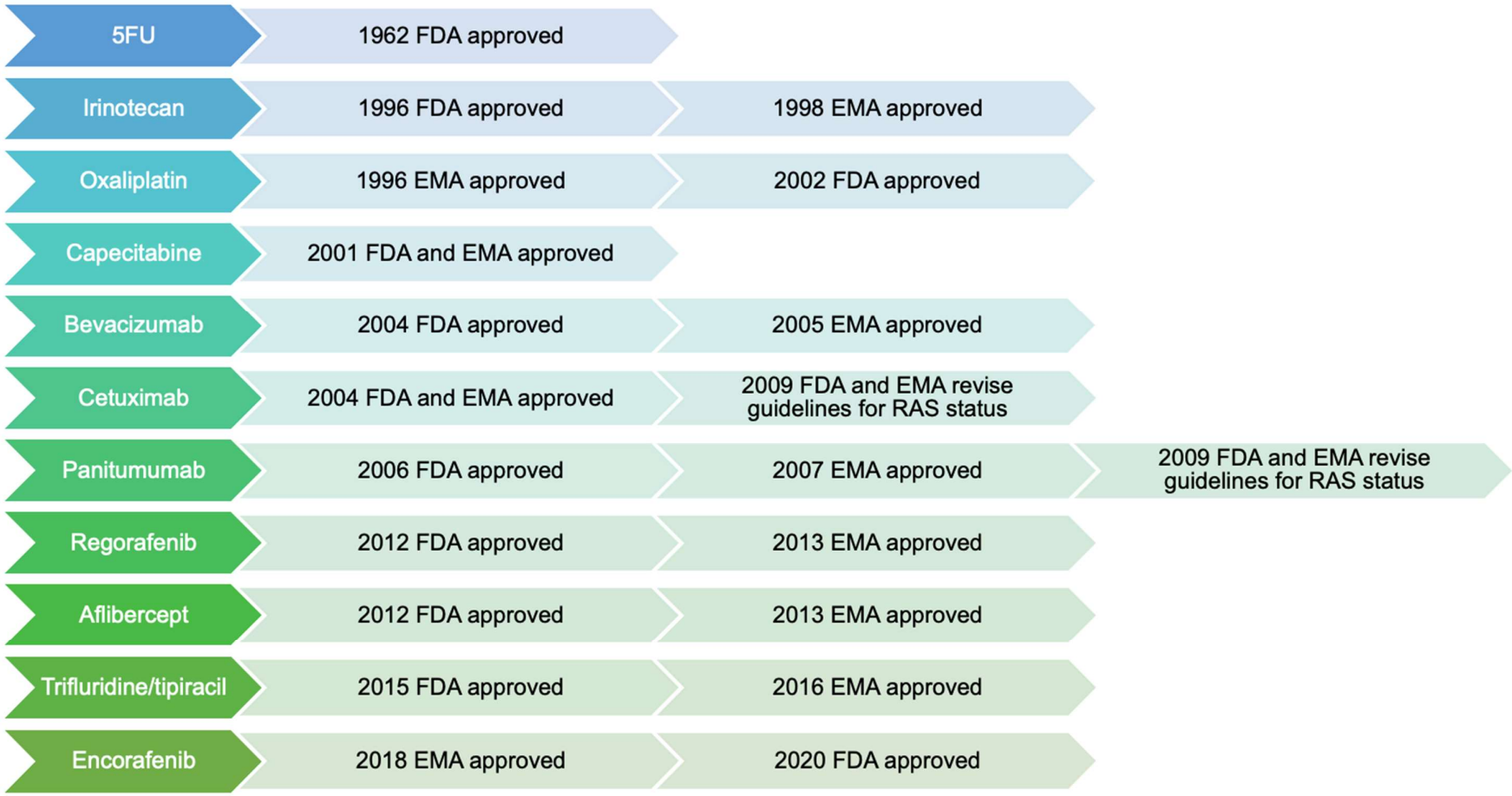

Figure 1.1 Treatments used for colorectal cancer. FDA = Food and drug administration, EMA = European Medicines Agency.

1.2.1.1 Fluorouracil (5FU)

5FU was the first chemotherapy developed for CRC and is still one of the most common chemotherapies used today (Gustavsson *et al*, 2015). For CRC, it is mainly used for advanced cancer or patients at high risk of recurrence (Blondy *et al*, 2020). 5FU is a fluoropyrimidine and works through multiple pathways (**Figure 1.2**; Longley *et al*, 2003; Francini *et al*, 1994). The main mechanism of effect is through inhibition of thymidylate synthase, which disrupts DNA synthesis and repair mechanisms and ultimately results in cell death**.** However, over 80% of administered 5FU is degraded by the enzyme dihydropyrimidine dehydrogenase (DPD) and excreted out in the urine (Miura *et al*, 2010). 5FU is often combined with folinic acid, also known as leucovorin. Folinic acid synergises with 5FU, significantly improving its efficacy (Piedbois *et al*,1992). It does so by stabilising the bond between the active metabolite and thymidylate synthase, allowing 5FU to remain in cancer cells longer (Moran and Keyomarsi, 1987).

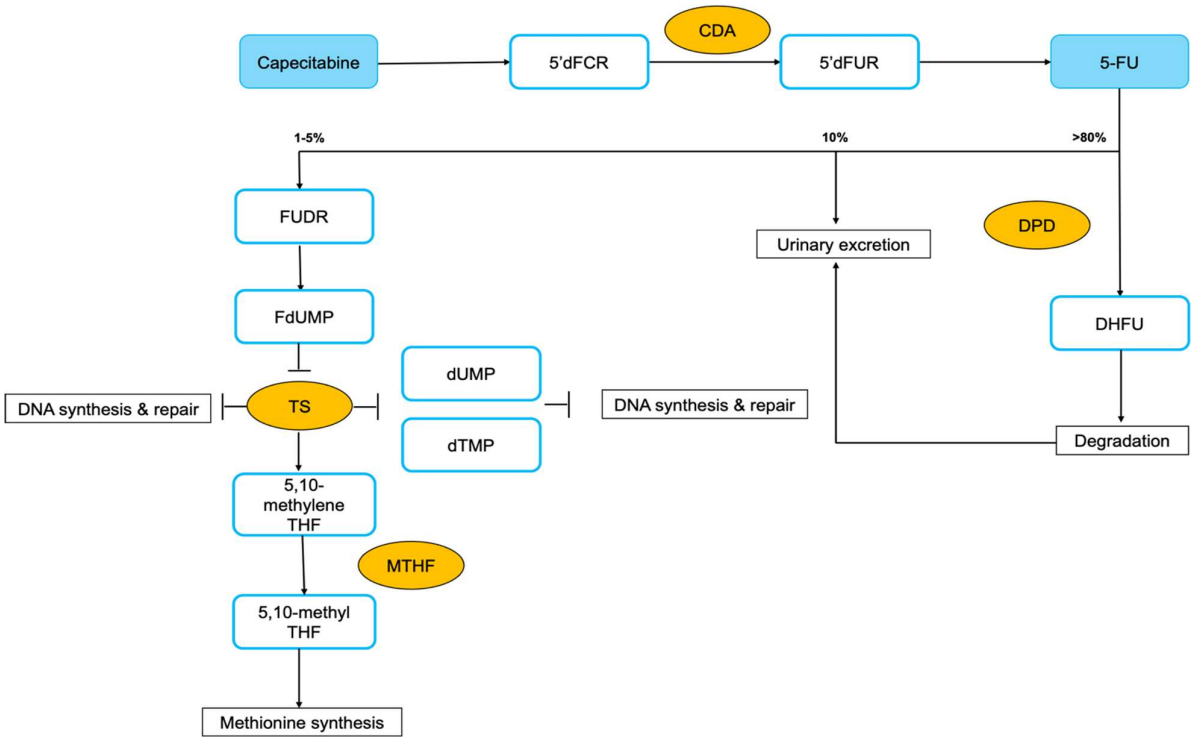

Figure 1.2 Simplified diagram of the 5FU metabolism pathway showing enzymes with suggested toxicity causing mutations. 5′dFCR= 5′deoxy-5-fluorocytidine; 5′dFUR= 5′deoxy-5-fluorouridine, DPD= Dihydropyrimidine dehydrogenase, DHFU= dihydro-fluorouracil, FUDR= fluorodeoxyuridine, FdUMP= fluoro-deoxyuridine-monophosphate, dTMP= thymidine monophosphate TS= Thymidylate synthase, THF= tetrahydrofolate, MTHF= 5-Methyltetrahydrofolate. Adapted from Escalante *et al* (2021). Capecitabine is converted into 5FU in the liver by the sequential action of carboxylesterase, CDA and thymidine phosphorylase. Approximately 80% of 5FU is catabolised into inactive metabolites by DPD. A further 10% is directly excreted out. Only 1% to 5% of 5FU is converted to active metabolites through nucleotide metabolic pathways. This leads to inhibition of TS which disrupts DNA synthesis and repair mechanisms. Another consequence of TS inhibition is the activation of enzymes involved in methionine synthesis, which increases 5FU activation in cancer cells. Mutations in CDA (CDA), *DPYD* (DPD) *TYMS* (TS), and *MTHFR* (MTHF) have been associated with 5FU toxicity (in yellow/gold).

1.2.1.2 Capecitabine

Capecitabine is a 5FU prodrug and was developed to be a more convenient alternative for patients, as it is administered orally (Van Cutsem *et al*, 2001). After several conversion steps, capecitabine is eventually broken down to 5FU and so has the same mechanism of action (**Figure 1.2**). Studies have shown that capecitabine is non-inferior to 5FU in terms of efficacy and survival (Hoff *et al*, 2001; Van Cutsem *et al*, 2001). However, some concerns remain about the toxicity profile of capecitabine compared to 5FU and therefore it has not become the primary treatment (Aguado *et al*, 2014). Currently, capecitabine is recommended for older adults or those with mobility problems as it can be administered at home rather than requiring frequent hospital visits (Aguado *et al*, 2014).

1.2.1.3 Oxaliplatin

Oxaliplatin is a third-generation platinum derivative chemotherapeutic agent. The main mechanism of effect is through the formation of DNA adducts, irreversibly damaging DNA and causing cell apoptosis (Alcindor and Beauger, 2011). Oxaliplatin is often combined with 5FU chemotherapies as studies have overwhelmingly shown better patient outcomes compared to 5FU monotherapy (Soulié *et al,* 1997; de Gramont *et al,* 2000; André *et al*, 2004). When combined with 5FU and folinic acid, the regimen is commonly called FOLFOX, which is the most common treatment for advanced CRC. There are several FOLFOX treatments such as FOLFOX-4 and FOLFOX-6 that differ in dosage and administration (Akdeniz *et al*, 2021). Oxaliplatin can also be combined with capecitabine, which is commonly referred to as XELOX or CAPOX.

1.2.1.4 Irinotecan

Irinotecan has a broad spectrum of activity, demonstrating strong anti-tumour activity against a wide range of tumours (Xu and Villalona-Calero, 2002; Bailly, 2019). It works by inhibiting topoisomerase I, a key enzyme in DNA replication, resulting in double strand DNA breakage and cancer cell death (Bailly, 2019). Irinotecan is normally only used for advanced stage CRC and in combination with other chemotherapeutics such as 5FU (Douillard *et al*, 2003). However, it can be used effectively as a monotherapy for second-line treatment (Van Cutsem and Peeters, 1998; Oostendorp *et al*, 2010).

1.2.1.5 Other chemotherapies used in the UK

Other treatments including Lonsurf (trifluridine and tipiracil) and raltitrexed have been licensed to treat CRC, but are not routinely used in the NHS (De Falco *et al*, 2019; Gustavsson *et al*, 2015). Reasons for this include expense, being less effective than current recommended therapies or they are not recommended as first-line treatment (NICE guidelines, 2022).

1.2.2 Targeted therapies

The development of targeted therapies has vastly increased the efficacy of CRC treatment (Xie *et al*, 2020). However, the combination of chemotherapy with more than one targeted therapy has not been shown to benefit patient outcomes but does increase toxicity incidence (Tol *et al*, 2009; Hecht *et al*, 2009). Therefore, current clinical practice is to administer only one targeted therapy. Several targeted therapies for CRC are available in the UK.

1.2.2.1 Cetuximab

Cetuximab is a monoclonal antibody first developed in 2004 that inhibits epidermal growth factor receptor (EGFR) activity (Cunningham *et al*, 2004). As EGFR is overexpressed in tumours and causes cell proliferation, this has proven to be an excellent target, with other EGFR inhibitors such as panitumumab subsequently developed (Hecht *et al*, 2004; Martinelli *et al,* 2007). However, there has been some conflicting evidence as to cetuximab's efficacy. In 2009, cetuximab was found to be only effective in patients without mutations in codons 12 and 13 of *KRAS or NRAS* (Van Cutsem *et al*, 2009). Around 30-40% of CRC patients carry these somatic mutations. Furthermore, a mutation in *BRAF* (V600E) was found to also affect the efficacy of cetuximab unless BRAF inhibitors were administered (Nicolantonio *et al,* 2008). Moreover, it has been suggested that perhaps only when combined with FOLFOX does cetuximab work at maximum efficacy (Bokemeyer *et* al, 2011; Qin *et al*, 2018).

1.2.2.2 Panitumumab

Panitumumab is an EGFR inhibitor similar to cetuximab in both mechanism and efficacy (Price *et al*, 2014; Modest *et al*; 2022). However, tumours that show resistance to cetuximab have been shown to still be sensitive to panitumumab in some cases (Montagut *et al*, 2012). Panitumumab can be administered effectively as monotherapy or in combination with chemotherapy, usually 5FU (Van Cutsem *et al,* 2007). Similarly, to cetuximab, panitumumab is ineffective for patients with *KRAS, NRAS* or *BRAF* mutations (Amado *et al*, 2008; Nicolantonio *et al*, 2008).

1.2.2.3 Bevacizumab

Bevacizumab is a monoclonal antibody that targets vascular endothelial growth factor (VEGF) (Hurwitz *et al*, 2004). Bevacizumab inhibits tumour growth by preventing the development of new blood vessels, which tumours rely on for sustenance and growth (Shih and Lindley, 2006). There are no known somatic mutations that alter effectiveness and is therefore suitable for patients with *KRAS, NRAS* or *BRAF* mutations (Price *et al*, 2011). Bevacizumab is also effective in combination with fluoropyrimidine based chemotherapies (Kabbinavar *et al*, 2005; Saltz *et al*, 2008).

1.2.3 UK approved treatments

All the therapies discussed in the previous sections have been approved for the treatment of CRC in the NHS, under NICE guidelines (NICE guidelines, 2012a; NICE guidelines, 2017). However, it is advised that cetuximab and panitumumab are only administered in patients without specified *KRAS*, *NRAS* or *BRAF* mutations, due to lack of efficacy. All of the other discussed therapies are suitable for patients with these mutations. It is also recommended that bevacizumab only be administered alongside non-oxaliplatin containing regimens.

1.3 CRC treatment toxicities

Toxicity, commonly known as side-effects or adverse events, are defined as unexpected medical problems, unfavourable changes in health or abnormal laboratory findings that occur during treatment with a drug or other therapy (National Cancer Institute, 2017). Some toxicities have short term acute effects whereas others remain after treatment has stopped (Andreyev *et al*, 2012). Toxicity adversely

affects a patient's quality of life and can be life threatening. Drug toxicity may result in treatment discontinuation or dose reduction (Blumenthal *et al*, 2019; Koedoot *et al*; 2003) thus significantly affecting the prospects of a cure (Huitema *et al*, 2002; Braun and Seymour, 2011). Most chemotherapeutic agents are associated with significant side-effects even when administered at normal therapeutic dosages. More than 40% of patients with solid tumours develop at least one severe toxicity during their treatment (Ingrand *et al*, 2020).

1.3.1 Toxicity classification system

During clinical trials, toxicity events are recorded using the common terminology criteria for adverse events (CTCAE), a descriptive terminology that provides a severity scale ranging from 0-5 for all possible toxicities (National Cancer Institute, 2017). Descriptions are provided for each grade to guide clinicians. In general, grade 0 represents absent toxicity, grade 1 represents mild toxicity, grade 2 represents moderate toxicity, grade 3 represents severe toxicity, grade 4 represents life-threatening toxicity and grade 5 represents death due the toxicity (**Table 1.1**). Several updates of CTCAE have been published with the latest version (v5.0) published in 2017.

Outside of clinical trials, toxicity events are not as well documented (Mandelblatt *et al*, 2015). In the UK there is no standard system for recording toxicity events and instead, a patient's GP or oncologist may record the event as they see fit using SNOMED CT (Wardle and Spencer, 2017).

Table 1.1 Common terminology criteria for adverse events grading scale

Grade	Descriptor	Grade guidelines
0	Absent	Absent
1	Mild	Asymptomatic symptoms, mild symptoms or diagnostic observations only. No clinical intervention was needed.
2	Moderate	Minimal local or non-invasive intervention was needed. Instrumental daily living activities such as grocery shopping may have been limited.
3	Severe	Medically significant but not life-threatening. Hospitalisation or pro-longed hospitalisation was required. May have been disabling and limited self-care daily living activities such as dressing or feeding self.
4	Life-threatening	Life-threatening consequences. Urgent or emergent intervention was needed.
5	Death	Death related to toxicity

The National Cancer Institute Common Terminology Criteria for Adverse Events is a descriptive terminology which is utilised for toxicity event recording during clinical trials. Not all grades are appropriate for all toxicities and therefore fewer grades may be described for some toxicities.

1.3.2 Clinical risk factors for toxicity to chemotherapeutics

Studies have shown that clinical factors can increase the risk of toxicity events. The largest risk factor is the combination of treatments administered (Braun and Seymour, 2011). Each drug has a unique toxicity profile with toxicities that are more likely to develop and that are often dose-limiting (**Table 1.2**).

Another key risk factor is treatment dosage with higher dosages causing more severe toxic events (Brock *et al*, 2021). Dosage can also have a cumulative effect so patients with longer or more frequent cycles are at an increased risk of both toxicity incidence and increased severity (Kerr *et al*, 2000; Bleiberg, 1998). The administration method can also play a role, with 5FU infusion causing less toxicity than bolus administration, as drug levels remain stable throughout treatment (Hansen *et al*, 1996).

On an individual level, there are further factors to consider. Age (Hurria *et al*, 2011), malnutrition (Seo *et al*, 2016) and having co-morbidities such as diabetes (Gu *et al*, 2021; Yokokawa *et al*, 2015), can increase risk significantly. Furthermore, ethnicity has been shown to affect the tolerability of 5FU, with East Asian patients having the lowest relative risk of toxicity and US patients having the highest relative risk (Haller *et al*, 2008).

Table 1.2 Common dose-limiting toxicities associated with colorectal cancer treatments

Treatment	Common dose-limiting toxicities	Frequency grade 1+	References
Oxaliplatin	Peripheral neuropathy Acute neuropathy	10-15% 89%	Saif and Reardon, 2005 Pachman et al, 2015
EGFR inhibitors - cetuximab and panitumumab	Acneiform rash	>60%	Pinto *et al*, 2011; Lacouture *et al*, 2018
5FU	Neutropenia	29-69%	Buroker *et al*, 1994; Garg *et al*, 2012
Capecitabine	Hand-foot syndrome	18- 77%	Kwakman *et al*, 2020; Cassidy *et al*, 2002; Tebbutt *et al*, 2010
Bevacizumab	Hypertension	23-44%	Ranpura *et al*, 2010; Kindler *et al*, 2005
Irinotecan	Diarrhoea Neutropenia	50-80% 63-77%	Stein *et al*, 2010 Park *et* al, 2019; Kuehr *et al*, 2004

1.3.3 Genetic risk factors for toxicity to chemotherapeutics

As each drug has a unique toxicity profile, they are likely also to have unique genetic variants associated with their profiles. There are two mechanistic categories to consider, the first is genetic variants that fall within drug metabolism genes. These patients present with severe toxicity that develops during the first couple of cycles (Pinto and Dolan, 2012). Typically, toxicity is widespread, and therefore these variants are not toxicity specific.

However, since most variants in metabolism genes are usually rare, these variants do not explain the observed inter-individual variation. Therefore, there are likely toxicity specific risk variants explaining at least of part this variability (Eichler *et al*, 2011). This would also explain why toxicity incidence rates vary across ethnicities (Haller *et al*, 2008). These variants would have no links to metabolism and therefore do not cause widespread toxicity, being only risk factors for individual toxicities or toxicities with shared causal pathways. These variants are harder to identify and validate since many toxicities have unknown or several mechanistic pathways (Stein *et al*, 2010; Vichaya *et al*, 2015; Pergolizzi *et al*, 2017). The majority of published studies have used a candidate gene approach targeted towards drug metabolism genes and therefore only a handful of variants outside these genes have been identified to date. Moreover, most of these are not validated, failed to reach significance after correction for multiple tests and/or had sample sizes less than 200 (Custodio *et al*, 2014; Won *et al*, 2012; Argyriou *et al*, 2013).

1.3.3.1 5FU

There have been several variants associated with toxicity to 5FU, the best documented of which lie within the gene encoding DPD, *DPYD* (Section 1.2.1.1 above). Genetic variants in *DPYD* can be life-threatening as some variants can completely inactivate DPD enzyme activity thereby preventing 5FU metabolism (Yen and McLeod, 2007). Approximately 3-5% of patients have at least one variant that causes partial or complete deficiency (Innocenti *et al*, 2020). Therefore, it is recommended that *DPYD* variants are genotyped before treatment, to guide dosage and treatment options. Currently, NHS patients are screened for four variants prior to 5FU therapy (**Table 1.3**). Patients who are heterozygous for any of these risk variants require a dose reduction between 25-75% and patients who are homozygous should not receive fluoropyrimidine chemotherapies (Morawska *et al*, 2018). There are a further 6 validated and 16 unvalidated variants in *DPYD* that have been associated with toxicity but either due to their limited predictive ability or conflicting evidence in other studies, they are not recommended for clinical use at present (**Table 1.3**).

Variants in Thymidylate Synthetase (*TYMS*) have also been associated with 5FU toxicity. While the biological relevance of *TYMS* is clear, the impact of noted variants is not (**Figure 1.2**). In the literature, two variants have been discussed extensively, a 28bp sequence occurring usually either as a double (2R) or triple (3R) tandem repeat (rs45445694) and a 6 base pair deletion (rs11280056). Both have been significantly associated with toxicity in some studies (Lecomte *et al*, 2004; Schwab *et al*, 2008; Castro-Rojas *et al*, 2017; Hamzic *et al*, 2020) but failed to replicate in several others (Martinez-Balibrea *et al*, 2010; Meulendijks *et al*, 2017; Braun *et al*,

2009; Sharma *et al*, 2008). However, it is unclear whether the causal gene is *TYMS* or the nearby gene enolase superfamily member 1 (*ENOSF1*), which may explain the lack of consistency in results. rs2612091 in *ENOSF1* has been associated with 5FU toxicity and is in partial linkage disequilibrium (LD) with both *TYMS* variants (Meulendijks *et al*, 2017; Hamzic *et al*, 2020). However, the mechanism of effect for *ENOSF1* is unclear, although it may regulate *TYMS* activity (Wu and Dolnick, 2003). Overall, further work is required to establish which gene is causal and the mechanism of effect before any of these variants could be implemented in the clinical setting.

Table 1.3 Validated variants associated with toxicities to chemotherapeutics used in the treatment of colorectal cancer

Treatment	Gene	SNP	Alternative names	Effect	EUR MAF	Original observation	Replication
5FU	*DPYD*	**rs3918290**	*2A, IVS14+1G>A	Splice variant	0.005	Salgueiro, 2004	Largillier, 2006
		rs55886062	*13A p.I560S	Missense	0.002	Morel, 2006	Lee, 2014
		rs67376798	p.D949V	Missense	0.007	Boisdron-Celle, 2007	Deenen, 2011
		rs75017182	c.1129-5923C>G	Intronic	0.02	Amstutz, 2009	Froehlich, 2015
		rs1801265	*9A p.C29R	Missense	0.23	Vreken, 1997	Joerger, 2015
		rs755416212	p.Arg235Gln	Missense	0.00002	van Kuilenburg, 2008	Ly, 2020
		rs2297595	p.M166V	Initiator Codon Variant	0.12	Gross, 2008	Falvella, 2015
		rs1801160	*6 p.V732I	Missense	0.05	Collie-Duguid, 2000	Kleibl, 2009
		rs115232898	p.Tyr186Cys	Missense	0.00004	Zaanan, 2014	Saif, 2014
		rs17376848	p.Phe632=	Synonymous	0.04	Kristensen, 2010	Falvella, 2015
	TYMS	rs45445694	TYMS 5'-UTR VNTR	Indel in promoter	-	Kristensen, 2010	Castro-Rojas, 2017
		rs11280056	TYMS 3'-UTR 6-bp ins	Indel in promoter	-	Rosmarin, 2015	Hazmic, 2020
	ENOSF1	rs2612091	c.742-227G>C	Intronic	0.44	Meulendijks, 2017	Hamzic, 2020
Capecitabine	*CDA*	rs2072671	p.Lys27Gln	Missense	0.25	García-González, 2015	Pellicer, 2017
	MTHFR	rs1801133	p.Ala263Val	Missense	0.27	Sharma, 2008	Gusella, 2009
		rs1801131	p.Glu470Ala	Missense	0.31	Sharma, 2008	Thomas, 2011
Oxaliplatin	*GSTP1*	rs1695	p.Ile105Val	Missense	0.37	Lecomte, 2006	Ruzzo, 2007
Bevacizumab	*KCNAB1*	rs6770663	c.276-10722A>G	Intronic	0.26	Quintanilha, 2022	Quintanilha, 2022
	SV2C	rs6453204	c.580+11045 A>G	Intronic	0.08	Schneider, 2014	Schneider, 2014
Irinotecan	*UGT1A1*	**rs3064744**	*28 g.234668881 TA[7]	Indel in promoter	0.39	Iyer, 2002	Innocenti, 2004
		rs10929302	*93 c. 3156G > A	Promoter	0.30	Innocenti, 2009	Hulshof, 2022
		rs4148323	*6 p.Gly71Arg	Missense	0.008	Han,2006	Cheng, 2014

EUR= , MAF= minor allele frequency. Indel = Insertion/deletion variant. Minor allele frequencies referenced from gnomAD v.2.1.1 (Europeans non-Finnish). Variants in bold are currently tested before therapy is administered.

1.3.3.2 Capecitabine

Variants in *DPYD* and *TYMS/ENOSF1* have also been associated with capecitabine toxicity (Rosmarin *et al*, 2015, Henricks *et al*, 2017). However, as there are additional conversion steps from capecitabine to 5FU, unique capecitabine specific markers have also been identified (**Table 1.3**). rs2072671 in cytidine deaminase (*CDA*) has been consistently associated with capecitabine toxicity although its clinical utility is limited due to its modest effect size of around 2 (García-González *et al*, 2015; Pellicer, 2017; Mattia *et al*, 2019). Several variants in drug transporter gene ATP Binding Cassette Subfamily B Member 1 (*ABCB1*) have been suggested to be associated with capecitabine toxicity but findings have been inconsistent across studies (Gonzalez-Haba 2010, Loganayagam *et al*, 2013; García-González *et al*, 2015; Mattia *et al*, 2019). Two variants in methylenetetrahydrofolate reductase (*MTHFR*) have also been associated with capecitabine toxicity but both routinely fail to associate with 5FU toxicity despite *MTHFR* being involved with 5FU and folate metabolism (Sharma *et al*, 2008; Loganayagam, 2013; Afzal *et al*, 2009; Mattia and Toffoli, 2009)**.** The biological reasoning for this is unclear (Loganayagam, 2013).

1.3.3.3 Oxaliplatin

Variants in key metabolism genes have been tested for association with oxaliplatin toxicity, but studies have routinely failed to identify any significant loci (**Figure 1.3**; Ruzzo *et al*, 2014; Varma *et al*, 2020; Park *et al*, 2022). Of the studies that have yielded results, all have proven contentious (Ye *et al*, 2013; Formica *et al*, 2017). For example, a variant in Glutathione S-Transferase π 1 (*GSTP1*) has been associated with peripheral neuropathy in several studies (**Table 1.3**; Lecomte *et al*, 2006; Ruzzo

et al, 2007; McLeod *et al*, 2010) but has been contested in numerous others (Gamelin *et al,* 2007; Inada *et al*, 2010; Peng *et al,* 2013).

1.3.3.4 Irinotecan

Variants in the irinotecan metabolism gene UDP glucuronosyltransferase family 1 member A1 (*UGT1A1*) have been associated with severe toxicity (**Table 1.3**; Karas and Innocenti, 2022). rs3064744 (*28) is recommended for clinical use as it is common in patients with African (43%) and European (39%) ancestries. Two other *UGT1A1* variants, rs10929302 (Innocenti *et al*, 2009; Hulshof *et al*, 2022) and rs4148323 (Han *et* al, 2006; Cheng *et al*, 2014) have also been associated with an increased risk of toxicity and could prove clinically useful. Other variants and genes have also been proposed, but without validation in independent cohorts (Innocenti *et al*, 2009; Han *et al*, 2013; Chen *et al*, 2015; Riera *et al*, 2020).

1.3.3.5 Cetuximab and panitumumab

There are no genetic markers with strong evidence associated with toxicity to cetuximab or panitumumab. Several studies have identified potential markers, but none of these have been validated in replication cohorts (Baas *et al*, 2018; Froelich *et al*, 2018).

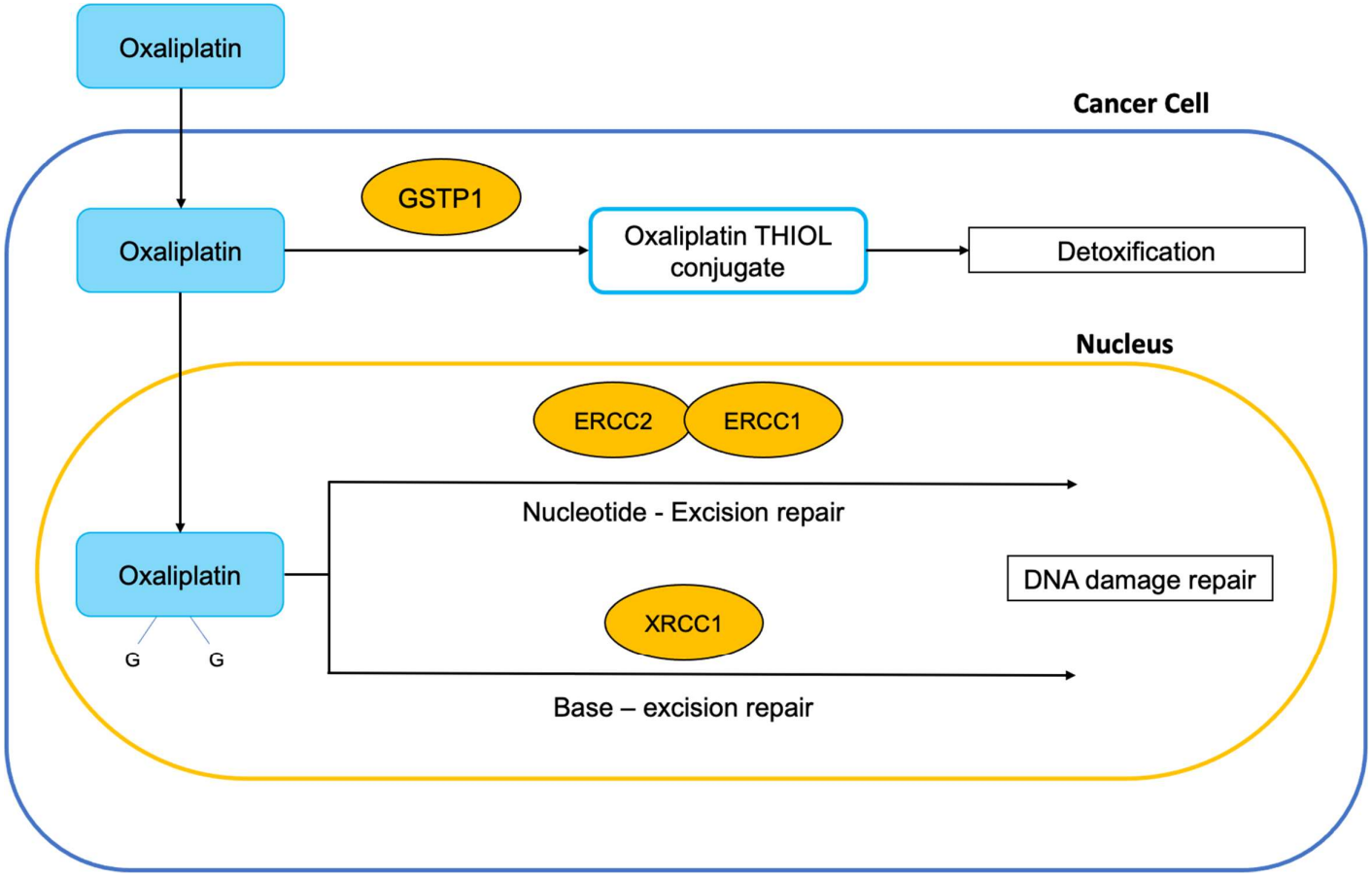

Figure 1.3 Simplified diagram of the oxaliplatin metabolism pathway showing enzymes with suggested toxicity causing mutations. Adapted from Escalante *et al* (2021). Some oxaliplatin is directly detoxified by *GSTP1* which is then eliminated. Oxaliplatin's main mechanism of cytotoxicity is through the formation of DNA adducts. In summary, oxaliplatin binds to the guanine and cytosine bases in DNA which creates DNA crosslinks which induces cell apoptosis. DNA-oxaliplatin adducts are then subsequently repaired through either the nucleotide-excision repair or base-excision repair pathways. Mutations in *GSTP1*, *ERCC2*, *ERCC1* and *XRCC1* (in yellow/gold) have been tested for association with oxaliplatin toxicity but have routinely proven non-significant .

1.3.3.6 Bevacizumab

Two variants have been associated with bevacizumab-induced hypertension which have been validated in independent cohorts (**Table 1.3**). One lies within potassium voltage-gated channel subfamily A regulatory beta subunit 1 (*KCNAB1*) (Quintanilha *et al*, 2022) and the other within synaptic vesicle glycoprotein 2C (*SV2C*) (Schneider *et al,* 2014). Neither of these genes are involved with bevacizumab metabolism but both are involved with biologically relevant pathways for hypertension (Li and Kroetz, 2018). Other variants have also been proposed for bevacizumab toxicity, but none have been validated in independent cohorts (Lambrechts *et al*, 2014; di Stefano *et* al, 2015; Li *et al*, 2018).

1.4 Genome-wide association studies (GWAS)

GWAS are a popular approach to identify single nucleotide polymorphisms (SNPs) associated with a phenotype of interest. GWAS function by testing for differences in allele frequency for each SNP across individuals that differ phenotypically but are similar in most other aspects including ancestry.

1.4.1 GWAS hypotheses

GWAS were first designed based on the common disease common variant hypothesis (**Figure 1.4**). This states that if a heritable disease is common in the population, then genetic contributors should also be common in the same population (Reich and Lander, 2001). However, common variants with large effect sizes would be subject to negative selection and so it is more likely variants have small effect sizes. Therefore, it is expected that most phenotypes are polygenic, with each variant contributing little individually, but collectively explaining phenotype heritability

(Bush and Moore, 2012). However, for toxicity-related traits, it has been hypothesised that negative selection may not be applicable unless causal variants are also associated with other harmful phenotypes. This is because exposure to chemotherapy is a recent phenomenon in evolution, so toxicity variants would not have been subjected to negative selection pressure (Maranville and Cox, 2015). Therefore, common toxicity-related variants may have large effect sizes.

There is also a competing hypothesis to consider, the common-disease rare variant hypothesis that states common diseases could be caused by multiplicity of individual rare genetic variants (Pritchard, 2001; Gorlov *et al*, 2008). Supporters have argued this hypothesis is more consistent with human pathobiology (Schork *et al*, 2009). Rare variants are more likely to be relatively new and have therefore not been subjected to strong negative selection or are rare because they have been selected against due to having a deleterious effect (Pritchard, 2001). In contrast, common variants are more likely to be older and by reaching a common frequency indicates they have not been subjected to negative selection (Schork *et al*, 2009). Therefore, some GWAS have now turned to look at low-frequency and rare variants, when common variants have failed to explain heritability (Bomba *et al*, 2017 Lettre, 2014).

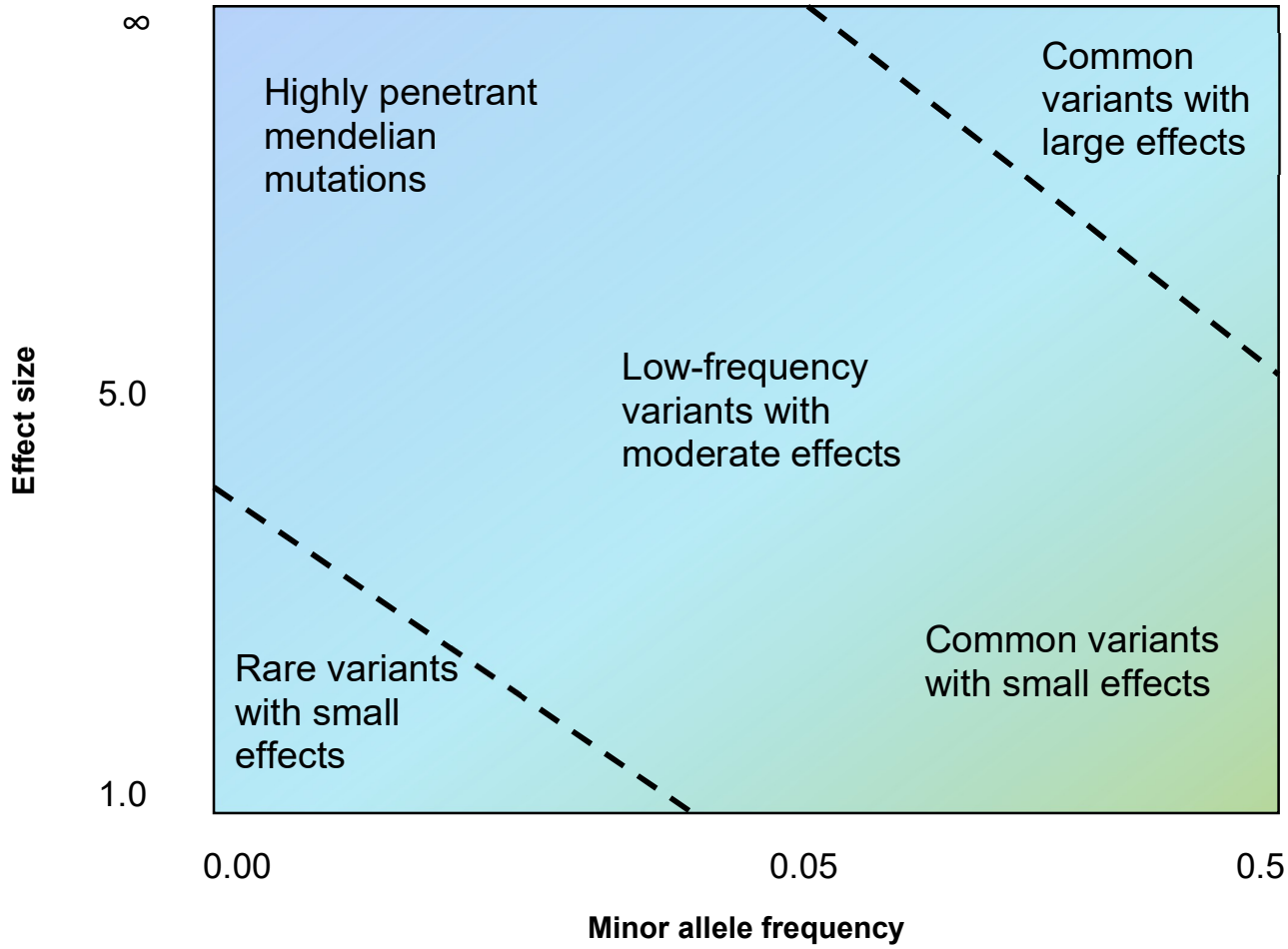

Figure 1.4 Relationship between the frequency of the variant minor allele and the size of the effect. The X-axis represents variant effect size, and the Y-axis represents variant minor allele frequency. Genome-wide association studies are effective at identifying common variants with small or large effect sizes. However, common variants with large effect sizes are rare, particularly for disease phenotypes. More studies are now investigating low-frequency variants that have historically been excluded from analysis. Rare variants with small effects are still virtually undetectable with current technology.

1.4.2 Single nucleotide polymorphisms

SNPs are defined as single base pair changes that occur in at least 1% of the general population and are one of the most common forms of genetic variation (Brookes, 1999). It is estimated that a person's genome differs from the reference genome at approximately 4.5 million sites, with most of these being common variants (100,000 genomes project consortium, 2015). Most known SNPs are classified as silent mutations, but some SNPs can have serious functional consequences. SNPs that have direct effects on protein structure can be classified as missense, frameshift, splicing, or nonsense mutations (Dobson *et* al, 2006; Shastry, 2009). SNPs can also have functional consequences that do not affect protein structure, including through gene expression, transcription factor binding sites and mRNA stability (Shastry, 2009).

There are also several different models of genetic inheritance for SNPs including dominant, recessive and additive models. Most association studies test for additive effects as an additive model can still capture non-additive variability with some accuracy, whereas other models cannot (Tsepilov *et al*, 2015). However, most additive GWAS will always have insufficient power to detect lower frequency SNPs with truly recessive effects (Guindo-Martínez *et al*, 2021).

1.4.3 Linkage disequilibrium

LD is the non-random association of alleles at different loci (Slatkin, 2008). High LD indicates that SNPs are usually inherited together and therefore are part of the same haploblock. LD between loci can be measured using two methods, R^2 and D'. Both can be useful indicators; R^2 measures the squared correlation between a pair of loci

and D' assesses the relationship between haplotype frequencies. LD is crucial to GWAS interpretation as the most associated SNP is often not causal (Schaid *et al*, 2018). The process of fine mapping can help to determine which SNPs have a high confidence of being causal. For GWAS studies, R^2 is the usual measure used (Bush and Moore, 2012).

1.4.4 Genotyping and imputation

SNP chip array genotyping is a cost-effective method to get large coverage of the genome (Verlouw *et al*, 2021). If genotyped SNPs are chosen wisely then additional information can be gained through imputation. Genomic imputation works by leveraging the power of LD between SNPs to infer the haplotypes of non-genotyped SNPs (**Figure 1.5**). The first step is haplotype phasing where the probability of each genotype is calculated per variant using a whole genome sequence (WGS) reference panel to infer genotypes (Howie *et al*, 2009; The Haplotype Reference Consortium, 2016). From these probabilities, imputation scores can be generated that represent the probability of the variant being called accurately. Only variants with high imputation scores will be used for analysis. However, no agreed threshold has been set although most studies use a cut off between 0.4-0.8 (Southam *et al,* 2011; Zheng *et al*, 2015; The Haplotype Reference Consortium, 2016).

Imputation has been shown to be very accurate for common variants, while historically being poor for rare variants. Therefore, most GWAS studies to date have excluded rare variants from analysis. However, newer larger reference panels have been shown to impute rare variants accurately and so rare variant analyses are likely to become commonplace in the future.

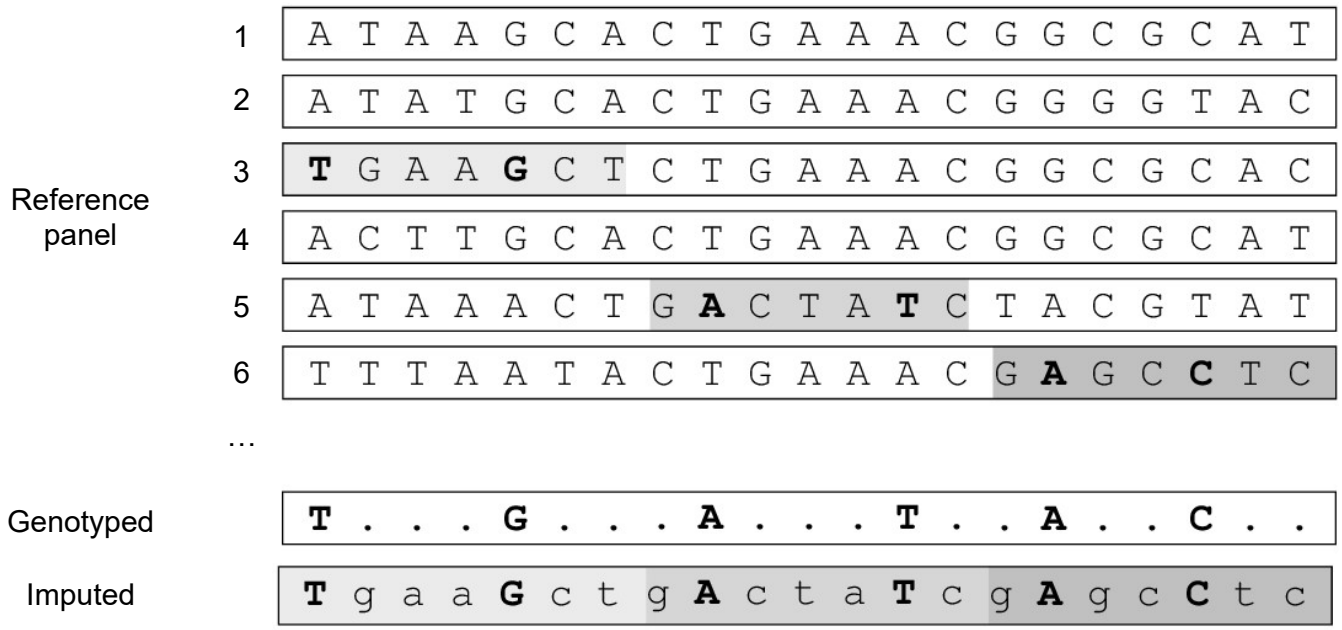

Figure 1.5 Genotype imputation methodology. Genotyped alleles from individual I are used to match segments from the reference panel. Non-genotyped sites can then be inferred using the matched segment. Adapted from Das (2017).

1.4.5 GWAS validation

Unfortunately, false positives are common in GWAS and therefore this must be considered when reporting results. Providing strong mechanistic or biological evidence can help validate the initial observation (König, 2011). To confirm the observation, the gold standard is to replicate significant GWAS results in an independent cohort, proving the result is replicable (Oetting *et al*, 2017). When choosing a replication cohort, it is important to consider heterogeneity between study samples, otherwise confounding may interfere with results (Liu *et al*, 2008)

1.4.6 Statistical power

Statistical power analyses are a critical first step of any statistical study (Hong and Park, 2012). Power analyses calculate the probability of finding an effect assuming that there is an effect to be found (Castelloe and O'Brien, 2001). Power is linked to several variables including sample size, case to control ratio, SNP effect size, SNP minor allele frequency (MAF) and the statistical significance threshold (Sham and Purcell, 2014). Changing one or more of these variables will influence power (**Figure 1.6**). For prospective studies, power calculations can be used to guide the design of the cohort and study. However, most GWAS are performed retrospectively, and power calculations are therefore used to guide analysis choices to maximise power (Sham and Purcell, 2014; Visscher *et al*, 2017).

For GWAS, due to the large number of SNPs tested, multiple testing burden must be considered. The *de facto* significance threshold for GWAS and use in power calculations is $P<5.0\text{x}10^{-8}$ which is an alpha of 0.05 corrected for 1,000,000 tests using Bonferroni (Risch and Merikangas, 1996; Bland and Altman, 1995). Although

the number of SNPs tested has increased since this threshold was set, over the years Bonferroni has been proposed to be too conservative due to LD and the threshold has therefore never been adjusted (Chen *et al*, 2021). A second threshold of suggestive significance has been established at $P<1.0x10^{-5}$ which indicates SNPs of potential interest. This was introduced due to the conservative nature of Bonferroni and because SNPs with low effect sizes or MAFs may be too underpowered to reach the significance threshold unless very large sample sizes are utilised (Lander and Kruglyak, 1995).

Power should also be calculated for validation cohorts to ensure that the SNP has sufficient power to replicate. While the original effect size of the SNP can be useful to guide this, it is also important to consider the winners curse phenomenon (Lohmueller *et al*, 2003). This states a variant effect size is likely overestimated in the original study and may be weaker in subsequent validation studies (Göring *et al*, 2001). Thus, the actual sample size of the validation cohort may need to be even larger than the initial estimate (Liu *et al*, 2008).

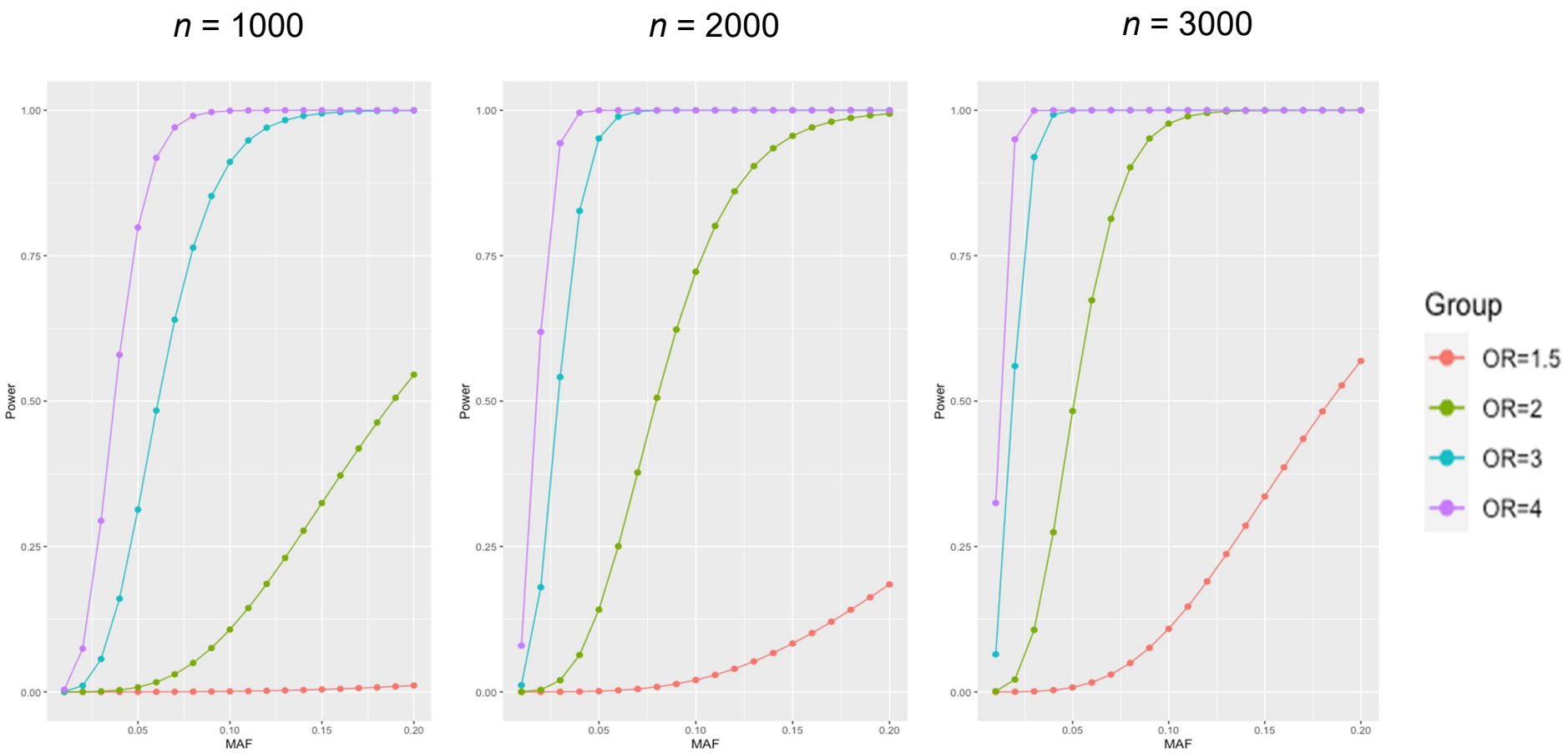

Figure 1.6 Genome-wide association power calculated based on a sample size of n with a case rate of 30%. The X-axis shows SNP minor allele frequency (MAF), and the Y-axis shows statistical power. Power plots are shown for studies with 1000, 2000 and 3000 participants (left to right). Power curves are shown for SNPs with odds ratios (OR) of 1.5, 2, 3 and 4. The disease model is assumed to be additive.

1.4.7 Covariates

Covariates are included in statistical analyses to reduce confounding and increase precision (McCaw *et al*, 2022). Precision will improve if the distribution of the phenotype varies across levels of the covariate by reducing residual variation (McCaw *et al*, 2022). The simplest and most common method to incorporate covariates is to include a linear or binary term for the covariate in the association model. However, when a covariate is not linearly associated with the phenotype, transformation of the covariate may be needed (Pain *et al*, 2018). In GWAS studies it is also possible to test for the effect of a covariate to see if a SNP of interest interacts with the covariate.

1.5 Other bioinformatic analyses

1.5.1 Molecular quantitative trait loci

In addition to SNPs directly affecting the phenotype, they can also shape phenotypes through gene expression, splicing or epigenetic changes (Qu *et al*, 2017). SNPs that are associated with this variation are called molecular quantitative trait loci (QTL). QTLs can be classified as *cis* or *trans* depending on their location of effect. In most cases, QLTs are as *cis* indicating that they regulate nearby genes (<1 Mb). In contrast, the rarer *trans* QLTs can regulate far away genes (>1 Mb) or genes on other chromosomes (Suzuki *et al*, 2021). Most QLTs are also tissue specific unless associated with house-keeping genes (Gerrits *et al*, 2009). The most common QTLs are expression QTLs (eQTL) and splice QTLs (sQTL). eQTLs are associated with gene expression and epigenetic changes. Studies suggest eQTLs are enriched in loci identified from GWAS which may help validate and explain the effect of non-coding loci (Kubota and Suyama, 2022). sQTLs are associated with the expression

of RNA isoforms from alternative splicing events (Garrido-Martín *et al*, 2021). In several diseases sQTLs have larger predicted effects than eQTLs (Farh *et al*, 2015; Yamaguchi *et al*, 2022).

However, some recent studies have shown that QTL variants are more likely to have smaller effect sizes and are unlikely to be within critical genes (Wang and Goldstein, 2020; Battle *et al*, 2014). This is because genes with key functional roles have been more conserved throughout evolution. Therefore, while QTLs help establish potential mechanistic pathways for intronic SNPs, not being a QTL should not exclude SNPs from further investigation.

1.5.2 Gene and gene set analyses

One key problem of GWAS is that SNPs often only explain a fraction of a phenotype's heritability, which can make them difficult to detect at genome-wide significant levels (Lee *et al,* 2011). Gene and gene set analyses are one avenue to gain additional insight from GWAS results while also placing the results in a broader biological context.

In gene analyses, single SNP association results are aggregated to the gene level. This makes it possible to detect genes where multiple weaker associations are present. There are various methods available which have different underlying assumptions (Purcell, 2007; Holmans *et al,* 2009; Segrè *et al*, 2010). Earlier software had several problems including being dependent on LD structure, computational demand and lack of interpretability. More recent algorithms such as MAGMA have

overcome these problems and are now frequently used post-GWAS (de Leeuw *et al*, 2015).

Gene set analyses also known as pathway analyses work by aggregating the gene analysis results into biological pathways. Significant gene sets indicate that genes within the set are enriched for association with the phenotype. Gene set analyses can lead to novel hypotheses for the biological mechanisms of diseases. Various programs are available however only INRICH and MAGMA have consistently shown good power (de Leeuw *et al*, 2015). This is important, as power strongly depends on the heritability of the phenotype with more heritable phenotypes being more weakly powered (de Leeuw, 2016).

1.5.3 *In silico* analyses

In silico analyses can be key in contextualising GWAS results by determining how SNPs can affect the phenotype of interest. Although SNPs that fall within exonic regions can affect protein structure directly, their effects can have varying impact. Several programs to predict the functional impact of coding variants include Combined Annotation Dependent Depletion (CADD) (Kircher *et al*, 2014), Polymorphism Phenotyping version 2 (PolyPhen2) (Adzhubei *et al*, 2010) and Sorting Intolerant from Tolerant (SIFT) (Sim *et al*, 2012). More difficult is determining the relevance of intronic SNPs or SNPs that fall outside protein coding genes. Looking at evolution conservation can help to indicate regions that may have important regulatory function (Huber *et al*, 2020). Intronic SNPs may also affect splicing which can be predicted with the use of tools such as Human Splicing Finder (Desmet *et al*, 2009) and SpliceDetector (Houreh *et al*, 2018).

1.6 Visualisation and plots

1.6.1 Manhattan plots

Manhattan plots are used to visualise and interpret the results of a GWAS (**Figure 1.7A**). A Manhattan plot shows the *P*-values of a GWAS ordered by chromosomal location. The Y axis is the $-\log_{10}$ (*P*-values), and the X axis is genomic location by chromosome and then chromosomal position (Turner, 2018). Genome-wide significance at $P<5.0\times10^{-8}$ and suggestive significance at $P<1.0\times10^{-5}$ are indicated with red and blue lines, respectively.

1.6.2 Quantile-quantile (QQ) plots

QQ plots graphically represent the deviation of the observed *P* values from *P*-values expected under the null hypothesis. The observed *P*-values are plotted in descending size order. If no genomic inflation is observed, the points will align to form a diagonal line from the bottom left corner to the upper right corner (**Figure 1.7A**). If over or under inflation is observed, points will deviate from the centre line. QQ plots are an easy method to determine if over or under inflation of test statistics is occurring. Over inflation usually indicates a population stratification issue and under inflation can indicate a lack of power (Reed *et al*, 2015).

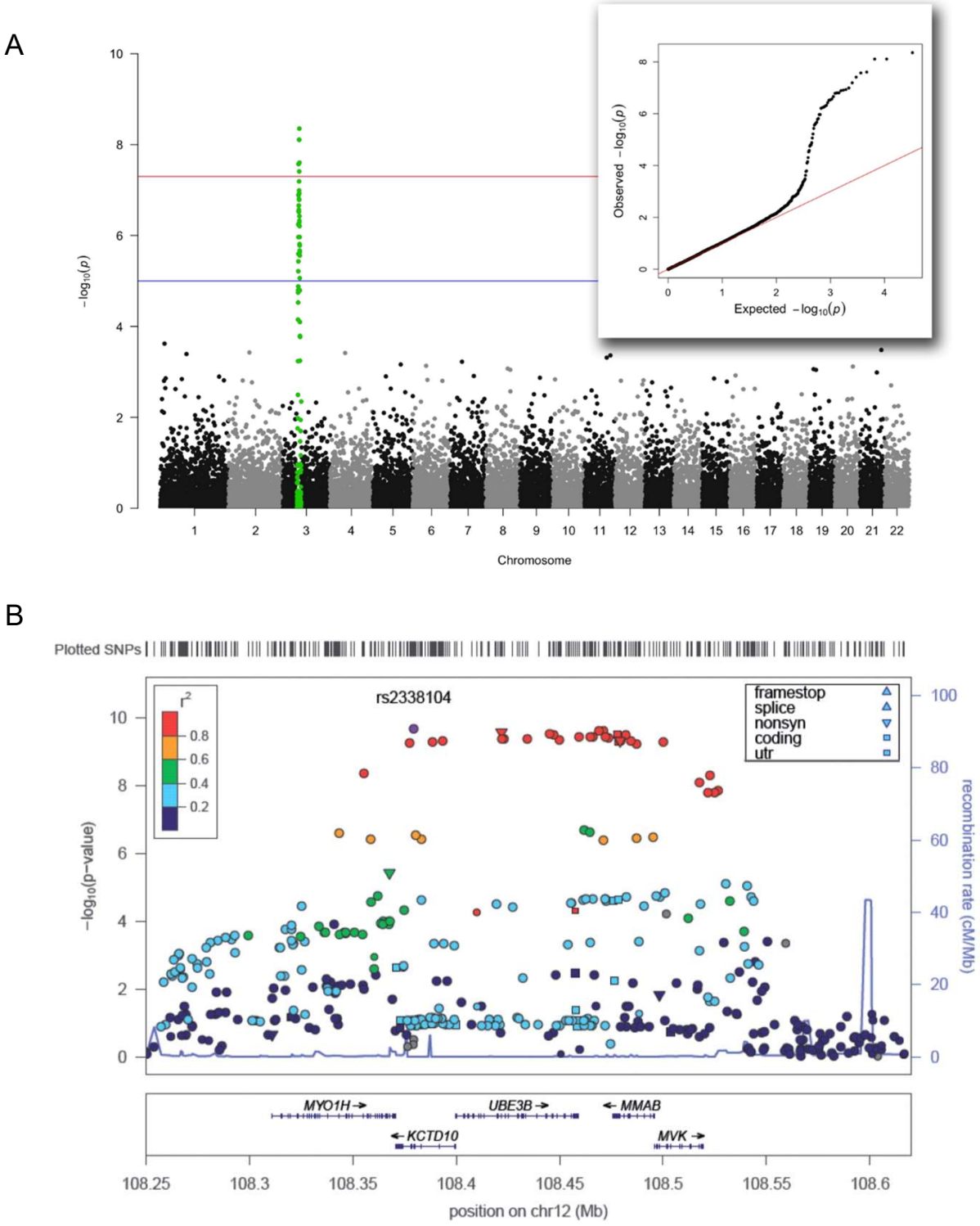

Figure 1.7 Visualisation of Genome-wide association study results. (A) Example Manhattan plot adapted from Turner (2018) and its QQ plot showing the expected versus observed *P*-values for SNPs. On the Manhattan plot, the red line

corresponds to a P=5.0x10^{-8} and the blue line P=1.0x10^{-5}. On the QQ plot, upward deviation from the diagonal red line indicates potential genomic inflation and correlates to high peaks in a Manhattan plot. **(B)** Example Locuszoom plot adapted from Pruim et al. (2010). The plot shows the results of the analysis for SNPs and recombination rates. −log10(P) (y-axis) of the SNPs are shown according to their chromosomal positions (x-axis). The sentinel SNP (purple) is labelled by its rsID. The colour intensity of each symbol reflects the extent of linkage disequilibrium (LD) with the sentinel SNP, deep blue (r2=0) through to dark red (r2=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 37 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. Functional annotation is shown using shapes as noted by the legend.

1.6.3 Locuszoom regional association plots

Regional plots are important to determine where SNPs lie within the genome and their relationship with nearby genes (**Figure 1.7B**). Several software is available including Locuszoom (Pruim *et al*, 2010), SNAP (Johnson *et al*, 2008) and CandiSNPer (Schmitt *et al*, 2010). Of these, Locuszoom provides extra features not available in other software including visualisation of LD structure, recombination peaks and the ability to incorporate finemaps or functional annotations (Pruim *et al*, 2010). These additional features can be vital for guiding downstream analyses. For example, it can be quickly determined using LD patterns how many independent association signals may be present at a locus.

1.7 Hypothesis and aims

The main hypothesis of this thesis is that there are germline SNPs associated with toxicity to chemotherapeutics that have not yet been discovered. I expect to identify SNPs that are associated with individual toxicities, some of which may be treatment specific. An analysis plan outlining the phenotypes and cohorts used throughout this thesis, is shown in **Table 1.4**.

Individual aims:

- Identify common toxicity SNPs that are associated with specific treatment combinations.
- Perform meta-analyses with QUASAR2 to identify SNPs associated with capecitabine toxicity.
- Identify low-frequency SNPs associated with 5FU toxicity.
- Validate any promising markers using external cohorts and *in silico* analyses.

Table 1.4 Simplified version of thesis analysis plan

<table>
<tr><th rowspan="2"></th><th rowspan="2">Phenotypes analysed</th><th colspan="4">COIN + COIN-B</th><th colspan="2">QUASAR2</th></tr>
<tr><th>FOLFOX</th><th>FOLFOX + cetuximab</th><th>XELOX</th><th>XELOX + cetuximab</th><th>Capecitabine</th><th>Capecitabine + bevacizumab</th></tr>
<tr><td>Chapter 3</td><td>Any-toxicity, diarrhoea, neutropenic sepsis, peripheral neuropathy, hand-foot syndrome, neutropenia, lethargy, stomatitis, nausea, vomiting, rash</td><td>GWAS</td><td>GWAS</td><td>GWAS</td><td>GWAS</td><td colspan="2"></td></tr>
<tr><td>Chapter 4</td><td>Diarrhoea, neutropenic sepsis, peripheral neuropathy, hand-foot syndrome, neutropenia, lethargy, stomatitis, nausea, vomiting, rash</td><td colspan="2">Meta-analysis (of 2 GWAS)</td><td colspan="2">Meta-analysis (of 2 GWAS)</td><td colspan="2"></td></tr>
<tr><td>Chapter 5</td><td>Diarrhoea, hand-foot syndrome, neutropenia, stomatitis, vomiting</td><td></td><td></td><td colspan="4">Meta-analysis (of 4 GWAS)</td></tr>
<tr><td>Chapter 6</td><td>Diarrhoea, neutropenic sepsis, peripheral neuropathy, hand-foot syndrome, neutropenia, lethargy, stomatitis, nausea, vomiting, rash</td><td colspan="4">GWAS</td><td colspan="2"></td></tr>
</table>

Each of the results chapters in this thesis (Chapter 3-6) presents results from either genome-wide association studies (GWAS) or GWAS meta-analyses. For boxes marked GWAS, all patients in the indicated groups were analysed as one cohort, whereas for boxes marked meta-analyses, GWAS were performed for each group and then meta-analysed together. For each chapter, greyed-out boxes indicate groups which were not used during the initial analysis phase. These groups were however, used for validation analyses.

2 Materials and Methods

2.1 My contribution and others' contributions

Genotyping, imputation and initial quality control (QC) of COIN (COntinuous vs INtermittent) and COIN-B was performed prior to this project. Quick and Simple and Reliable trial (QUASAR2) data is held by Claire Palles, Birmingham University and summary statistic data from their analyses was provided for this thesis. The UK Biobank and Genomic England (GEL) cohorts had standard QC performed by their respective companies. Christopher Wills (Cardiff University) performed additional QC of the UK Biobank data. Further QC measures imposed on each dataset are detailed in the relevant results chapter's materials and methods sections. All other analyses were performed by myself unless otherwise stated.

2.2 Datasets used in this thesis

2.2.1 COIN and COIN-B clinical trials

2.2.1.1 COIN trial design and aims

COIN was a Cancer Research UK and MRC funded phase III clinical trial (ISRCTN27286448). COIN aimed to determine (i) if there was a significant difference in patient outcomes between those who received continuous or intermittent therapies and (ii) if the addition of cetuximab had a significant effect on patient outcomes for those receiving continuous therapies (Maughan *et al*, 2011; Adams *et al*, 2011).

Recruitment ran between 2005 and 2008 across 111 hospitals in the UK and Ireland. Inclusion criteria included having written informed consent, being at least 18 years old, having a histologically confirmed adenocarcinoma of the colorectum, inoperable

metastatic or locoregional measurable disease (RECIST v1.0, Therasse *et al*, 2000), being chemotherapy naive for metastatic disease, a WHO performance status 0–2 and having good end-organ function (Adams *et al*, 2011). Patients were excluded if they had previous or present malignant disease, had uncontrolled medical comorbidity that may interfere with treatment or assessment, had known brain metastases or if they had ever been administered oxaliplatin.

In total, 2,245 patients were recruited and randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy (n=815), continuous chemotherapy with cetuximab (n=815), or intermittent chemotherapy (n=815) (**Figure 2.1**). Two chemotherapies XELOX (XEL= capecitabine, OX= oxaliplatin) and FOLFOX (FOL= folinic acid, F= fluorouracil, OX= oxaliplatin) were utilised in the trial and administered based on the choice of the patient and doctor. In total, approximately one third chose FOLFOX and two thirds XELOX**.** For patients in arms A and B the trial ceased when one of the following occurred, disease progression, severe toxicity or patient choice (Adams *et al*, 2011).

The results showed non-inferiority in patient outcomes between those receiving intermittent chemotherapy compared to those receiving continuous therapy. Overall survival (OS) was an average of 14.4 months for intermittent *versus* 15.8 months for continuous cetuximab, but the difference was not significant (Hazard Ratio [HR]=1.08, 95% Confidence interval [95%CI]=0.97-1.21, *P* not reported). There was also no evidence of cetuximab having a significant improvement on patient outcomes when comparing all patients or those with *KRAS* wild type tumours (Maughan *et al*, 2011).

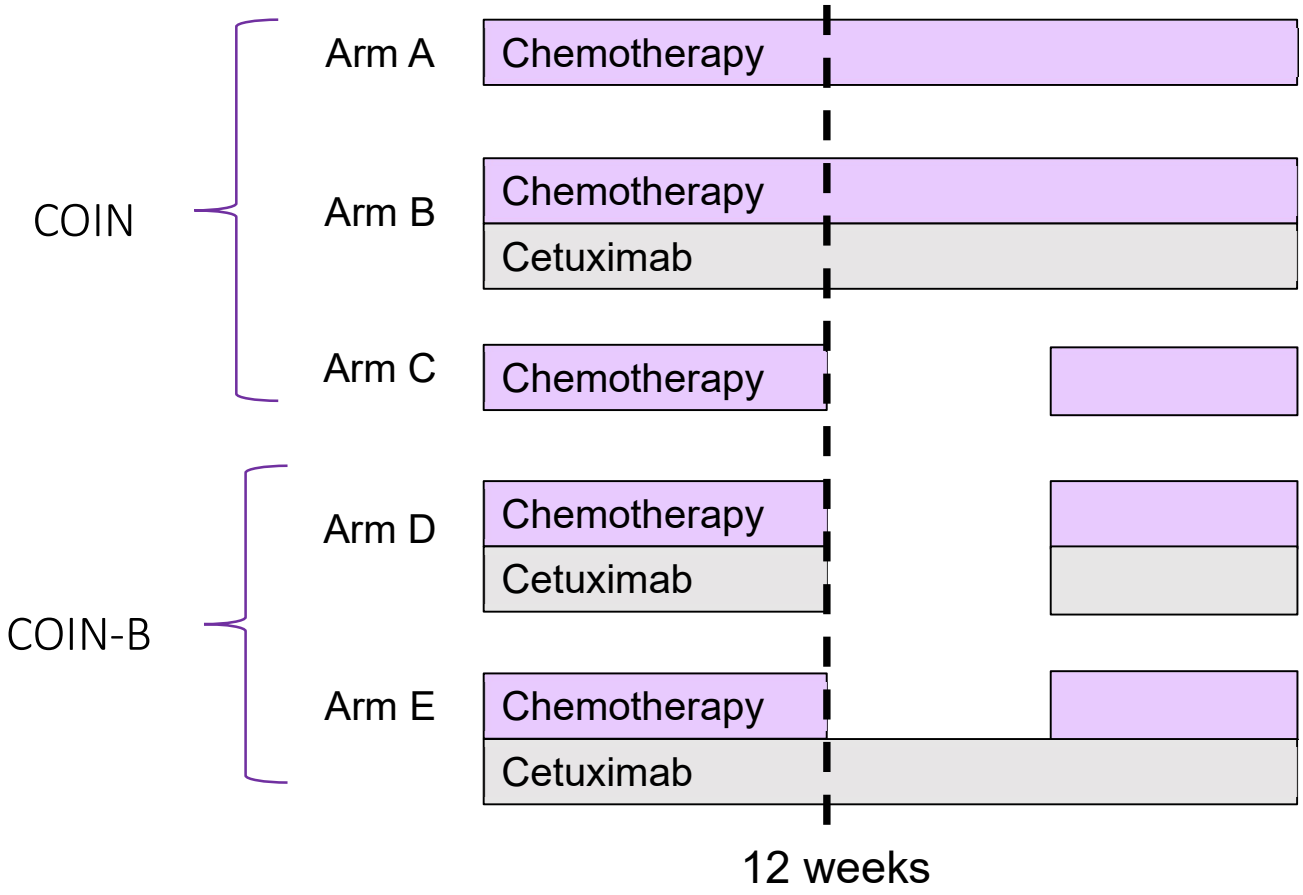

Figure 2.1 COIN and COIN-B trial design. In COIN, patients were randomised 1:1:1 to receive continuous oxaliplatin and fluoropyrimidine chemotherapy (n=815), continuous chemotherapy with cetuximab (n=815), or intermittent chemotherapy (n=815). In COIN-B, patients were randomised 1:1 to receive intermittent chemotherapy and cetuximab (n=112) or intermittent chemotherapy and continuous cetuximab (n=114). For patients in intermittent therapy arms, treatment was given for 12 weeks and then resumed when disease progression occurred, and 12 more weeks of treatment was administered.

2.2.1.2 COIN-B trial design and aims

COIN-B was a follow-on study and a phase II clinical trial (ISRCTN38375681). COIN-B aimed to determine if there was a significant difference in patient outcomes between patients receiving continuous cetuximab or intermittent cetuximab (Wasan *et al*, 2014).

Recruitment ran between 2007 and 2010 across 30 hospitals in the UK and one in Cyprus. In May 2008, COIN-B was halted and re-designed after data indicated that *KRAS* mutations were predictive of resistance to EGFR targeted therapies. Tumoral *KRAS* mutation status of already recruited patients was assessed and the protocol for future recruitment amended to include screening for *KRAS* mutation status. Only the participants with *KRAS* wild type tumours were recruited following reactivation of the trial. Other inclusion criteria included having written informed consent, being at least 18 years old, having a histologically confirmed adenocarcinoma of the colorectum, inoperable metastatic or locoregional measurable disease (RECIST v1.0), a WHO performance status 0–2 and having good end-organ function (Eisenhauer *et al*, 2009). Patients were excluded if they had any uncontrolled medical comorbidity that may interfere with treatment or had a history of cancer or brain metastases.

In total, 226 patients were recruited and randomised 1:1 to receive intermittent chemotherapy and cetuximab (n=112) or intermittent chemotherapy and continuous cetuximab (n=114). All patients received FOLFOX. All patients received treatment for 12 weeks followed by a break in chemotherapy (and cetuximab for Arm D patients) until RECIST disease progression occurred, at which point they received 12

additional weeks of chemotherapy. Patients in Arm D received cetuximab continuously throughout the trial.

Results indicated there was better outcomes for patients with *KRAS* wild type tumours when treated with continuous compared to intermittent cetuximab. Progression free survival was an average of 3.1 months (95%CI=2.8-4.7) for intermittent *versus* 5.8 months (95%CI=4.9-8.6) for continuous cetuximab. Similarly, failure free survival was an average of 16.8 months (95%CI=14.5-22.6) for intermittent *versus* 22.2 months (95%CI=18.4-28.9) for continuous cetuximab.

2.2.1.3 Cohort demographics

Given their overlapping treatment regimes, COIN and COIN-B were treated as one cohort for my genetic analyses. In total, 2,671 patients (mean age at randomisation of 62 years, range 18-87, 36% female) with metastatic or locally advanced CRC were recruited. For the first 12 weeks, treatments were identical in all patients apart from the choice of fluoropyrimidine in COIN (n=1,603, 60% received XELOX and n=1,068, 40% received FOLFOX) together with the randomisation of ± cetuximab (n=1,041, 39% received cetuximab).

2.2.1.4 Genotyping and QC

Blood DNA samples were prepared from 2,244 of the 2,671 patients and were genotyped using Affymetrix Axiom Arrays according to the manufacturer's recommendations (Affymetrix, Santa Clara, CA 95051, USA) (Al-Tassan *et al*, 2015).

Genotyping QC was tested using duplicate DNA samples with a concordance rate of >99%. Individuals were excluded from analysis if they failed one or more of the following checks: overall genotyping SNP rate <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 Principal Component (PCA)-based analysis (n=130). After QC, SNP genotypes were available for 1,950 patients.

Phasing of genotypes was performed using SHAPEIT and prediction of untyped SNPs was carried out using IMPUTE2 (v2.3.0) using the 1000 Genomes Project as a reference panel. As part of this thesis, I applied stricter QC filters before analysis. SNPs with imputation scores < 0.8, genotyping rates <95% and deviation from Hardy-Weinberg equilibrium (HWE) > $1x10^{-6}$ were excluded from analyses.

2.2.1.5 Toxicities to chemotherapeutics

Assessment of toxicities was performed at 12 weeks, since 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. Toxicities assessed were diarrhoea, neutropenic sepsis, peripheral neuropathy, hand-foot syndrome (HFS), neutropenia, lethargy, stomatitis, nausea,

vomiting and rash graded by critical adverse events as per the CTCAE (v4.0) with the highest grade noted within the first 12 weeks of treatment. Note, for HFS and nausea, the maximum possible grade on the CTCAE scale is 3 (severe). For all other toxicities the maximum grade is 5 (death).

Out of 1,950 patients with genotyping data, 150 did not have toxicity data and these were excluded leaving 1,800 for analyses. The cohort demographics for the 1800 patients are shown in **Table 2.1** (and analysed by treatment). The frequencies of toxicities in COIN and COIN-B are shown in **Table 2.2**.

Table 2.1 Clinicopathological data for patients from COIN and COIN-B by treatment received

		FOLFOX	FOLFOX + cetuximab	XELOX	XELOX + cetuximab	Difference between groups *P*-value
N		385	360	707	348	
Age at diagnosis	Mean (SD)	62.3 (9.4)	62.2 (9.8)	62.9 (9.5)	62.4 (9.2)	0.68
	Range	22.0-87.0	25.0-81.0	18.0-83.0	36.0-82.0	
Sex	Female	136 (35)	142 (39)	239 (34)	106 (30)	0.08
	Male	249 (65)	218 (61)	468 (66)	242 (70)	
WHO Performance status	0	166 (43)	166 (46)	343 (49)	182 (52)	0.29
	1	186 (48)	169 (47)	314 (44)	145 (42)	
	2	33 (9)	25 (7)	50 (7)	21 (6)	
Primary site of tumour	Colon	221 (57)	206 (57)	364 (51)	179 (51)	0.11
	Other	164 (43)	154 (43)	343 (49)	169 (49)	
Number of metastatic sites	0	2 (1)	1 (0)	6 (1)	4 (1)	0.76
	1	139 (36)	137 (38)	231 (33)	128 (37)	
	2	143 (37)	133 (37)	295 (42)	133 (38)	
	3	82 (21)	68 (19)	144 (20)	69 (20)	
	4	17 (4)	18 (5)	28 (4)	13 (4)	
	5	2 (1)	3 (1)	3 (0)	1 (0)	
Liver metastases	Yes	299 (78)	265 (74)	525 (74)	263 (76)	0.55
	No	86 (22)	95 (26)	182 (26)	85 (24)	
Lung metastases	Yes	159 (41)	134 (37)	300 (42)	142 (41)	0.44
	No	226 (59)	226 (63)	407 (58)	206 (59)	
Peritoneal metastases	Yes	61 (16)	58 (16)	99 (14)	55 (16)	0.74
	No	324 (84)	302 (84)	608 (86)	293 (84)	
Nodal metastases	Yes	170 (44)	169 (47)	349 (49)	147 (42)	0.13
	No	215 (56)	191 (53)	358 (51)	201 (58)	
Other metastases	Yes	60 (16)	68 (19)	107 (15)	51 (15)	0.37
	No	325 (84)	292 (81)	600 (85)	297 (85)	
White blood cell count	Mean (SD)	9.2 (3.3)	8.5 (3.1)	9.0 (4.4)	8.9 (3.5)	0.08
	Range	3.3-33.1	3.2-27.7	3.1-90.0	3.4-33.0	
Creatinine clearance	Mean (SD)	89.2 (27.7)	89.0 (27.5)	87.8 (27.6)	88.6 (29.6)	0.85
	Range	47.0-261.0	40.0-223.0	50.0-290.0	38.0-270.0	
Alkaline phosphatase	Mean (SD)	208.8 (195.3)	174.9 (149.1)	192.0 (179.1)	180.7 (176.3)	0.05
	Range	12.0-1456.0	33.0-1173.0	18.0-1452.0	27.0-1497.0	

Platelets	Mean (SD) Range	364.3 (136.0) 128.0-869.0	335.1 (120.7) 115.0-809.0	347.7 (125.8) 92.0-999.0	358.8 (140.7) 120.0-848.0	0.01*
Prior adjuvant chemotherapy	Yes No	85 (22) 300 (78)	82 (23) 278 (77)	192 (27) 515 (73)	87 (25) 261 (75)	0.22

For categorical variables, the number and percentage of patients in each category is stated and the *P*-value calculated using the Chi-squared test. For continuous variables, the mean, standard deviation (SD) and range are stated, and the *P*-value calculated using one-way ANOVA. *Not significant after correction for multiple testing.

Table 2.2 Patients from COIN and COIN-B with grade 2-5 CTCAE toxicities at 12 weeks

	FOLFOX treated		XELOX treated	
		+ cetuximab		+ cetuximab
	n=385 (%)	n=360 (%)	n=707 (%)	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 (6)	39 (11)	5 (0.7)	1 (0.3)
Peripheral neuropathy	43 (11)	30 (8)	110 (16)	44 (13)
HFS	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)

Percentage of patients in parentheses. HFS - Hand-foot syndrome.

2.2.2 QUASAR2

2.2.2.1 Trial design and aims

QUASAR2 (ISRCTN45133151) was an international phase III clinical trial to assess if the addition of bevacizumab to capecitabine therapy improved outcomes in patients with stage III or high-risk stage II CRC (Kerr *et al*, 2016) (**Figure 2.2**). The primary endpoint was 3-year disease free survival.

Patients were recruited between 2005-2010 across 170 hospitals in 7 countries (Australia, Austria, Czech Republic, New Zealand, Serbia, Slovenia and the UK). Recruitment criteria included being age 18 years or older, a WHO performance status of 0-1, histologically proven stage III or high-risk stage II CRC, primary resection 4-10 weeks before randomisation and life expectancy of at least 5 years. Exclusion criteria included having a history of cancer, inflammatory bowel disease, active peptic ulcer in the previous 2 years, or both; lack of physical integrity of the upper gastrointestinal tract, malabsorption syndrome, or inability to take oral medication; creatinine clearance <30 mL/min; absolute neutrophil count lower than $1.5x10^9$ cells per L; platelet count lower than $100x10^9$ cells per L; total bilirubin concentration higher than 1.5 times the upper limit of normal; alanine aminotransferase, aspartate aminotransferase, or alkaline phosphatase concentration greater than 2.5 times the upper limit of normal; proteinuria worse than 500 mg per 24 h; previous chemotherapy, immunotherapy, or infradiaphragmatic radiotherapy or the need for radiotherapy to these sites expected within the next 12 months; use of any investigational drug, agent, or procedure within 4 weeks of randomisation; chronic use of full-dose anticoagulants, high-dose aspirin, antiplatelet drugs, or known bleeding diathesis; concomitant treatment with sorivudine or its

chemically related analogues; history of uncontrolled seizures, central nervous system disorders or psychiatric disorders that precluded giving informed consent or interfered with adherence to oral drug intake; clinically important cardiac disease; known coagulopathy; known allergy to Chinese hamster ovary cell proteins; pregnancy, lactation, or no use of contraception in premenopausal women (Kerr *et al*, 2016). All participants provided their written consent and separate consent was obtained for the use of tumour and blood samples for further analyses.

In total 1,941 patients were recruited and randomised (1:1) to receive either capecitabine alone (n=968) or capecitabine and bevacizumab (n=973). Patients in the capecitabine and bevacizumab arm had the same dosing schedule for capecitabine as the capecitabine alone group, but with the addition of bevacizumab. Bevacizumab was administered for 8 extra cycles after capecitabine treatment for all patients ended.

The results indicated no significant difference in patient outcomes between those receiving capecitabine alone compared to those receiving bevacizumab + capecitabine (Kerr *et al*, 2016). Furthermore, the addition of bevacizumab was associated with increased toxicity (221 grade 3+ toxicity events in the capecitabine arm *versus* 350 grade 3+ toxicity events in the bevacizumab + capecitabine arm) so the authors recommended that this treatment combination should not be used in the treatment of CRC.

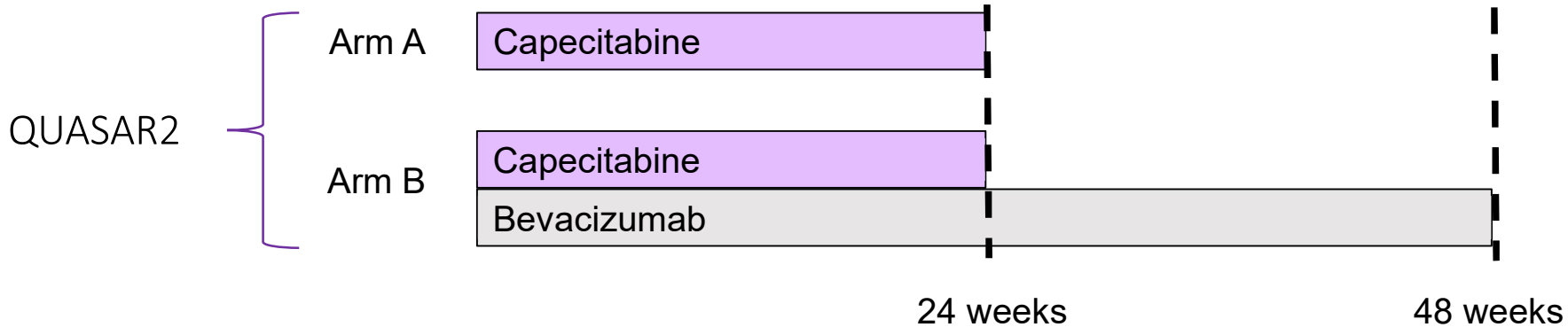

Figure 2.2 QUASAR2 trial design. Patients were recruited and randomised (1:1) to receive either capecitabine alone (n=968) or capecitabine and bevacizumab (n=973). Capecitabine was administered in 3-week cycles for a total of 24 weeks. Bevacizumab was also administered in Arm B every 3 weeks for a total of 48 weeks.

2.2.2.2 Patient demographics

On average QUASAR2 patients had a mean age at randomisation of 65 years (interquartile range 58-71) and 43% were female. 61% of all patients had stage III CRC and 38% had stage II CRC. Treatment was identical in all patients except for the randomisation of ± bevacizumab (n=973, 50% received bevacizumab).

2.2.2.3 Genotyping

The full genotyping methodology can be found in Rosmarin *et al* (2014). In brief, blood DNA samples were available for 1,119 patients and genotyped using Illumina genome-wide SNP panels (Human Hap 370, Human Hap 610 or Human Omni 2.5). Samples were excluded when there were poor genotyping call rates (<95%) or non-Caucasian ancestry by PCA-based analysis. Genotypes were imputed with IMPUTE (v2) using the 1000 genomes as a reference panel. A panel of 196 UK CRC patients with whole genome sequence data was used to check the accuracy of the imputation. SNPs were excluded if they had imputation scores <0.8 or a missingness rate above 10%.

2.2.2.4 Toxicities to chemotherapeutics

Toxicity events were recorded every 3 weeks for all patients receiving treatment, from enrolment until 30 days after the last dose of any study treatment was administered. Toxicities were recorded using the CTCAE (v3.0) grading system. In total, 930 patients had both genotyping and toxicity data available. The most common grade 2+ toxicities at 12 weeks were HFS (n=376), diarrhoea (n=199) and stomatitis (n=69).

2.2.3 UK Biobank

The UK Biobank data was accessed under application number 65833, Project entitled 'Investigating genetic & clinicopathological factors underlying risk, survival and toxicity to treatment in patients with cancer and population controls.

2.2.3.1 Cohort design

The UK Biobank is an open access population-based cohort of around 500,000 participants ranging in age between 40-69 years at recruitment (Bycroft *et al*, 2018). During recruitment participants signed consent forms including for follow-up through linkage to their health-related records. The initial assessment focused on lifestyle, health and socio-economic factors and took place between 2006 and 2010 across 22 assessment centres in the UK. Physical measurements were also recorded, and blood, saliva and urine samples were collected. Later, additional tests were added to the initial assessment including, eye measurements, electrocardiographs and hearing tests. Some participants have had up to 3 follow-on assessments which are repeats of the initial assessment, to track changes over time. Additional questionnaires continue to be sent out to expand the collection of phenotypic data.

2.2.3.2 Genotyping

The full genotyping methodology is available in Bycroft *et al*, 2018. In brief, DNA was extracted from blood DNA samples collected at the initial assessment. Genome-wide genotyping for 450,000 participants was performed using the UK Biobank Axiom Array which was carried out by the Affymetrix Research Services Laboratory. A further 50,000 samples were genotyped using the UK BiLEVE array. Approximately 850,000 variants were directly measured, and a further > 90 million variants were

imputed using the Haplotype Reference Consortium and UK10K + 1000 Genomes as reference panels.

Prior to analyses, participants were excluded if there was evidence of non-white European ancestry by PCA-based analysis, or if they had more than 10% of SNPs missing. Relatedness was calculated to identify strongly related pairs of individuals. Out of the pair, only one participant was kept in the cohort - if one had a history of cancer, they were kept, otherwise the choice was random. This left approximately 336,000 participants for use in analysis. SNPs were filtered to remove multiallelic variants, those with imputation scores <0.8, MAFs of < 0.01, missingness rates > 5% or with a deviation from HWE > $1x10^{-6}$.

2.2.3.3 Phenotypic data

2.2.3.3.1 International classification of diseases dataset

Data field 41270 contains the international classification of diseases version 10 (ICD-10) diagnosis codes extracted from participants hospital inpatient records. Inpatients were classified as people admitted to a hospital who occupied a bed for any length of time. This data field contains all ICD-10 diagnosis codes, that were recorded as either the primary or secondary diagnosis, for each participant. However, each UK nation has its own coding guidelines which could impact how hospital admissions were coded. Records date back to 1997 for England, 1998 for Wales and 1981 for Scotland. Participants never admitted to the hospital within the period covered are still present in the data field but with no recorded values.

Data field 41262 links to data field 41270 and contains the corresponding date for when each ICD-10 code was recorded. If an ICD-10 code appeared in the participants hospital records multiple times, only the earliest date was extracted.

2.2.3.3.2 Self-reported illness dataset

Data field 20002 contains self-reported non-cancer illnesses. Participants were interviewed by a nurse practitioner during recruitment and asked to provide a detailed medical history. When participants were uncertain of an exact diagnosis, they described the symptoms to the interviewer who attempted to identify the illness. If the illness could not be identified, then the interviewer recorded the information as free text. These were subsequently examined by a doctor to be classified if possible. Any free-text descriptions that could not be classified with a very high probability were marked as “unclassifiable”.

2.2.3.3.3 Toxicity to chemotherapeutics

Data on toxicities to drugs or chemotherapeutics was not available in the UK Biobank. Instead, phenotype data was extracted from the ICD-10 dataset, and I identified participants given chemotherapy using the classification of operations and procedures (OPCS4) dataset. Participants with codes X35.2, X37.3, X38.4 and X70-74 were included in the analyses as these codes indicate the participant was given chemotherapy. Cases were classified as participants that experienced a phenotypic event within 1 or 3 months of being given chemotherapy. Controls were classified as participants administered chemotherapy without an event.

2.2.3.3.4 Blood assays

Category 100081 contains results from haematological assays for basophils, eosinophils, monocytes, neutrophils, haematocrit, haemoglobin, red blood cells, lymphocytes, white Blood Cells, platelets and reticulocytes. Assays were performed using the whole blood sample collected during the initial assessment. 477,193 participants have data available for analysis.

2.2.4 Genomics England

Genomics England (GEL) data was accessed under application number 681, Project entitled 'Investigating genetic & clinicopathological factors underlying risk, survival and toxicity in patients with cancer and immune disorders.

2.2.4.1 Cohort design

Recruitment for the study ran between 2014 and 2018, and in total approximately 90,000 participants have been recruited and processed (Smedley *et al*, 2021). The study consisted of two arms, rare diseases and cancer. Participants were recruited by a range of clinical practitioners including doctors, clinical nurses and geneticists. NHS consultants could also nominate eligible persons for consideration. Children and adults of all ages were eligible for recruitment.

Participants were recruited for the rare disease arm if they had a disorder affecting < 1 in 2,000 persons, which was likely to have a single gene or oligogenic cause and had not received a genomic diagnosis (Smedley *et al*, 2021). Where possible, parents or other closely related family members of the participants were also recruited. The youngest affected person recruited in a family was assigned as the proband. Approximately 34,000 probands and 38,000 family members were

recruited. Baseline clinical data was recorded using human phenotype ontology terms for each participant.

Participants were recruited for the cancer arm if they had one of the eligible cancers listed on the GEL website. Participants were referred to the 100,000 genomes project by a member of their clinical team. Approximately 17,000 participants were recruited. Treatment plans were unchanged by any work performed by GEL. Cancer information such as staging and location was recorded at recruitment.

For all participants, whole blood DNA samples were collected during recruitment. All participants also had their health records (previous and future) linked to the study for use in analyses.

2.2.4.2 Sequencing

The full sequencing methodology is available in Smedley *et al* (2021) and on the GEL website (https://re-docs.genomicsengland.co.uk/sample_qc/). In brief, DNA was extracted from whole blood at the National Institutes for Health Research BioResource Laboratory in Cambridge. All genomes were sequenced with 150bp paired-end reads in a single lane of an Illumina HiSeq X instrument and uniformly processed on the Illumina North Star Version 4 Whole Genome Sequencing Workflow (NSV4, v2.6.53.23). Alignments had to cover at least 95% of the genome at 15X or above with well mapped reads or the sample was failed. Any samples with high levels of cross-contamination, mismatches with the declared gender, or for which consent had been withdrawn, were also excluded.

Prior to analyses, participants were also excluded if they were a relative of a proband or if there was evidence of non-white European ancestry by PCA-based analysis, or if the participant reported non-white ancestry.

2.2.4.3 Phenotypic data

2.2.4.3.1 International classification of diseases dataset

GEL extracted ICD-10 diagnosis codes from hospital inpatient records in a method similar to the UK Biobank. Inpatients were classified as people admitted to a hospital who occupied a bed for any length of time. GEL assembled a dataset of ICD-10 diagnoses codes, that were recorded as either the primary or secondary diagnosis, for each participant. Each record also contains the corresponding date of when each ICD-10 code was recorded. When an ICD-10 code appeared in the participants records multiple times, all the dates were listed.

2.2.4.3.2 Toxicity to chemotherapeutics

Data on toxicities to drugs or chemotherapeutics was not available in GEL. Instead, phenotype data was extracted from the hospital inpatient records dataset, and I identified participants given chemotherapy using the OPCS4 dataset. Participants with codes X35.2, X37.3, X38.4 and X70-74 were included in the analyses. Cases and controls were classified as described for the UK Biobank.

2.3 Hardware

All work presented in this thesis was performed using an Apple MacBook Pro (15 inch, 2019, Intel core I9, 32GB RAM, 1TB) using the operating system macOS Monterey purchased from Apple Incorporated (Cupertino, USA). Advanced Research

Computing at Cardiff (ARCCA) granted access to Cardiff University's high-performance computer (HPC) Hawk which was used via command line-based remote access for processes that required intensive computation such as analysing UK Biobank data.

2.4 Software

2.4.1 Downloaded software

Analyses were performed using software designed for statistical and genetic analyses. Plink (v1.9 and v2.0) were used for file conversion of genotype files and for statistical analyses and was downloaded from http://pngu.mgh.harvard.edu/purcell/plink/ . SNPTEST (v2) was used for calculating imputation scores and downloaded from https://www.well.ox.ac.uk/~gav/snptest/. IMPUTE2 (v2.3.2) was used to update imputation for areas of interest and was downloaded from https://mathgen.stats.ox.ac.uk/impute/impute_v2.html .MAGMA (v1.07 and v1.08) was used for gene and gene set analyses and were downloaded from https://ctg.cncr.nl/software/magma. The java version of Locuszoom (v0.13.2) was used for regional association plots and was downloaded from https://github.com/statgen/locuszoom/. PAINTOR (v3.1) was used for fine mapping and was downloaded from https://github.com/gkichaev/PAINTOR_V3.0. PHESANT was used for analysing blood assay phenotypes for individual SNPs using UK Biobank data and was downloaded from https://github.com/MRCIEU/PHESANT. R (v3.5.2) downloaded from http://www.r-project.org was used in conjunction with RStudio (v1.4.1106) downloaded from https://www.rstudio.com/, for survival analyses, data manipulation and visualisation of results.

2.4.2 R Packages

Packages for R were downloaded from Comprehensive R Archive Network (CRAN, https://cran.r-project.org/) and GitHub (https://github.com/). R packages used in this thesis are listed in **Table 2.3**. Some of the listed packages automatically install package dependencies which may not be listed here.

2.5 Statistical analyses

2.5.1 Power considerations

Power to detect odds ratios (OR) was calculated using the package genpwr in R. The calculation accounts for a *P*-value threshold, defined power and SNP MAF (1%, 5% or 20%). The result indicates the odds ratio needed for a SNP to pass the defined *P*-value threshold.

2.5.2 Genome-wide association analyses

GWAS were run under a univariate additive model in Plink (v1.9, Purcell *et al*, 2007) and results were plotted in R studio using qqman (Turner *et al*, 2018). A logistic regression method was chosen. SNPs that showed an association at $P<1.0x10^{-5}$ (suggestive of significance) were selected for independent validation.

Table 2.3 R Packages used in this thesis

Package	Use	References
devtools	Development tool for R packages	Wickham *et al* (2021a)
dplyr	Tool for data manipulation	Wickham *et al* (2021b)
plyr	Tools for splitting, applying and combing data	Wickham *et al* (2011)
genpwr	Power calculations for genetic models	Moore and Jacobson (2021)
ggplot2	Data visualisation	Wickham (2016)
qqman	Visualisation for Genome-wide association studies	Turner (2018)
RColorBrewer	Colour pallets for visualisation	Neuwirth (2014)
survival	Survival analysis	Therneau (2000)
survminer	Survival analysis	Kassambara et al (2021)
data.table	Tool for working with tabular data	Dowle *et al* (2021)
readr	Tool for reading data into R	Wickham *et al* (2020)
png	Tool for exporting figures as png files	Urbanek (2013)
stats	Package containing functions for statisitical calculations	R Core Team (2021)
tibble	Tool for creating dataframes	Müller *et al* (2021)
tidyr	Package for manipulating datasets	Wickham *et al* (2021c)

2.5.3 MAGMA gene and gene set analyses

MAGMA (de Leeuw *et al*, 2015) was used for gene and gene set analyses using data files from the NCBI 37.3 gene definitions and ~8,500 predefined gene sets. MAGMA uses an F-test to compute gene *P*-value. The snpwise multiple regression model was chosen to ensure that linkage disequilibrium (LD) between SNPs was fully accounted for during analysis. Gene analyses were run under a univariate model imposing a Bonferroni corrected significance threshold of $P=2.5x10^{-6}$.

Gene set analysis was run under the competitive model (de Leeuw et al., 2015). The competitive model tests if the genes in a gene set are more strongly associated with the phenotype than other genes and corrects for this. MAGMA also adds gene size and gene density as covariates to the regression models. Significance was set at a Bonferroni corrected significance threshold of $P=5.8x10^{-6}$.

2.5.4 Other bioinformatic analyses

2.5.4.1 Fine mapping

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 (Kichaev *et al*, 2014), which employs a Bayesian permutation method incorporating ENCODE and FANTOM5 functional annotations. Credible sets of causal SNPs were assembled for 95% coverage.

2.5.4.2 Blood assay analyses using UK Biobank

The software PHESANT (Millard *et al*, 2018) was used for single SNP association analyses against blood assay data in the UK Biobank. All analyses were adjusted for

sex, age at recruitment and genotyping chip. All blood assay data was continuous and analysed using linear regression except for where limited variation in the data was detected (where more than 20% of the samples have the same value). In these cases, the software defaults to using an ordered logistic model instead.

2.5.4.3 The genotype-tissue expression project database

The genotype-tissue expression project database (GTEx) is a publicly available resource which can be accessed at https://gtexportal.org/home. GTEx is comprised of samples from 838 donors, aged 20–79 years old with 67% being male. 84.6% of donors were white, 12.9% African-American, 1.3% Asian and 0.2% American Indian with the remaining donors having unknown heritage. The methodology of DNA sequencing can be found at https://gtexportal.org/home/documentationPage.

The GTEx database was used to identify eQTLs and sQTLs for SNPs of interest. Significance for tissue association was set at $P<1.0x10^{-3}$ (i.e. Bonferroni correction for 49 tissues [0.05/49]).

2.5.4.4 LocusZoom

Locuszoom produces high quality plots for visualising GWAS results. Locuszoom can be accessed on the web at http://locuszoom.org/ or be downloaded as a Java version for enhanced functions. The 95% credible sets from fine mapping are shown on the plot as squares. LD of nearby variants was calculated in relation to the lead SNP.

2.6 Study design

All the analyses performed in this thesis are retrospective with pre-determined sample sizes due to the recruitment of patients to the COIN and COIN-B trials. Reported *P*-values are uncorrected and two tailed unless otherwise stated. All analyses are reported in accordance with Strengthening the Reporting of Genetic Association Studies (STREGA) guidelines (Little *et al*, 2009).

3 Genome-wide association studies of toxicity to oxaliplatin and fluoropyrimidine chemotherapy with or without cetuximab

3.1 Introduction

3.1.1 XELOX and FOLFOX

The combination of fluoropyrimidine and oxaliplatin is a common first line treatment for many cancers including CRC (Stein and Arnold, 2012). XELOX is an oral fluoropyrimidine containing regimen with similar efficacy to FOLFOX, an intravenous fluoropyrimidine containing regimen, and both chemotherapies frequently cause toxicities. The Food and Drug Administration (2015) reported that 94% of patients administered FOLFOX and 96% administered XELOX will experience at least one toxicity over the course of their treatment. However, the two chemotherapies have different toxicity profiles (Ducreux *et al*, 2011; Guo *et al*, 2016). XELOX often causes gastrointestinal symptoms and HFS, whereas FOLFOX tends to affect immunity. This difference in profiles suggests that the underlying mechanisms of toxicities may be different.

3.1.2 Cetuximab

Cetuximab is also used in the treatment of CRC and commonly causes skin rash (Petrelli *et al*, 2013). The addition of cetuximab to chemotherapies also exacerbates the toxicity caused by the chemotherapy backbone (Huang *et al*, 2016).

3.1.3 Genetic variants associated with toxicities to fluoropyrimidines

Since there is significant inter-individual variation in chemotherapy-related toxicity (**Chapter 2, Table 2.2**), 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 *DPYD* are well established to be associated with severe toxicities to 5FU (Schwab *et al*, 2008; Henricks *et al*, 2018). While the role of common genetic variation is less clear, others have shown that common variants in *DPYD* also appear to affect toxicity (Gonzalez and Fernandez-Salguaro, 1995; Wei *et al*, 1996; Madi *et al*, 2018).

3.1.4 Aims

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 replication. To address such limitations, I have analysed GWAS data on 1,800 patients with advanced CRC treated with oxaliplatin and fluoropyrimidine chemotherapy ± cetuximab, with replication in independent patient groups.

3.2 Materials and methods

3.2.1 Patients and samples

Toxicity data and SNP genotypes were available for 1,800 patients from the COIN and COIN-B clinical trials after QC (Chapter 2, Sections 2.2.1.4 and 2.2.1.5) (**Figure 3.1**). Additional imputation was performed using IMPUTE2 for an 800Mb region surrounding *MROH5* (to provide better SNP coverage) using the phase 3 1000 Genome Project as reference. I restricted the analysis to directly typed SNPs and imputed SNPs with imputation scores of ≥0.8, a HWE of $\geq 1.0 \times 10^{-6}$ and a MAF of ≥0.05.

3.2.2 Clinical endpoints assessed and power considerations

The primary endpoint assessed was any toxicity, with patients that experienced any grade 2+ toxicity classified as a case (Chapter 2, Section 2.2.1.5). Secondary endpoints were individual toxicities: diarrhoea, neutropenic sepsis, peripheral neuropathy, HFS, neutropenia, lethargy, stomatitis, nausea, vomiting and rash. Patients with toxicities graded 2-5 were compared against those graded 0-1.

Logistic regression models were used to determine if the chemotherapy regimen and cetuximab administration affected toxicity occurrence. Power to detect toxicity effect sizes was calculated, based upon 70% power, a standard GWAS significance of $P=5.0 \times 10^{-8}$ and SNPs with MAFs of 0.20.

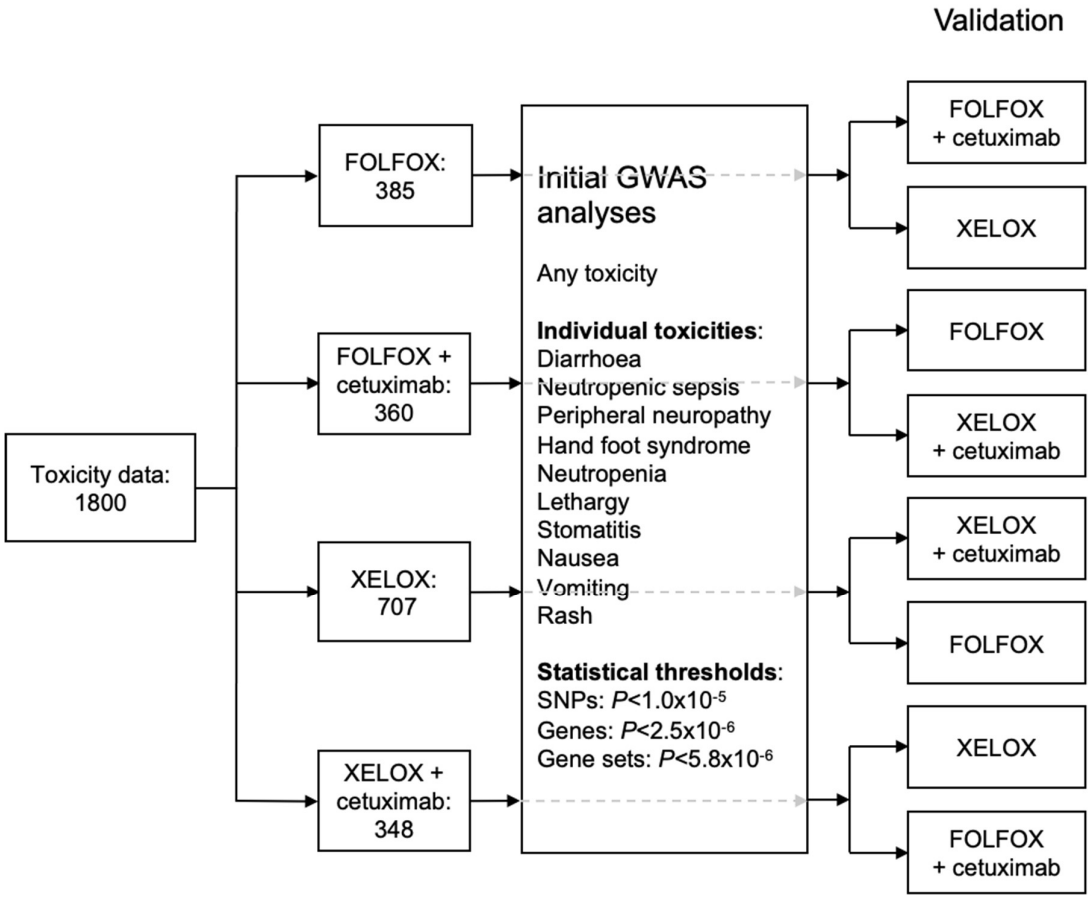

Figure 3.1 CONSORT diagram of the analysis strategy. The 1,800 patients were segregated into groups according to chemotherapy regimen and cetuximab status (385 patients received FOLFOX, 360 FOLFOX + cetuximab, 707 XELOX and 348 XELOX + cetuximab). I conducted genome-wide association studies for any toxicity and ten individual toxicities together with gene and gene set analyses. SNPs, genes and gene sets that reached genome-wide or suggestive significance were independently replicated in the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status, and, the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status.

3.2.3 GWAS analyses

Patients from COIN and COIN-B were analysed for associated genetic biomarkers after segregating by chemotherapy regimen and cetuximab status; 385 patients had FOLFOX, 360 had FOLFOX + cetuximab, 707 had XELOX and 348 had XELOX + cetuximab (**Figure 3.1;** Chapter 1, **Table 1.4**). The number of cases and controls for each treatment group are shown in Chapter 2 (**Table 2.2)**. GWAS were run under a univariate additive model in Plink (v1.9) and results were plotted in R studio using qqman (Chapter 2, Section 2.5.2). A logistic regression method was chosen. SNPs that showed an association at $P<1.0x10^{-5}$ (suggestive of significance) were selected for independent replication.

3.2.4 MAGMA gene and gene set analyses

MAGMA (de Leeuw *et al*, 2015) was used for gene and gene set analyses (Chapter 2 Section 2.5.3). Gene analyses were run under a snpwise univariate model imposing a Bonferroni corrected significance threshold of $P=2.5x10^{-6}$ (**Figure 3.1**). Gene set analyses were run under a competitive model with a corrected significance threshold of $P=5.8x10^{-6}$ (**Figure 3.1**).

3.2.5 Replication analyses

SNPs, genes and gene sets that reached genome-wide or suggestive significance in the GWAS analyses, were independently replicated in: (i) the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status, and (ii) the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status (**Figure 3.1**). For example, a SNP identified from the group receiving FOLFOX was replicated in those receiving FOLFOX + cetuximab and in

those receiving XELOX. A SNP identified from the group receiving XELOX was replicated in those receiving XELOX + cetuximab and in those receiving FOLFOX. A SNP identified from the group receiving FOLFOX + cetuximab was replicated in those receiving FOLFOX and in those receiving XELOX + cetuximab. A SNP identified from the group receiving XELOX + cetuximab was replicated in those receiving XELOX and in those receiving FOLFOX + cetuximab (**Figure 3.1**). I considered a nominally significant threshold of $P<0.05$ as evidence for replication. There was >85% power to detect the initially observed odds ratios for each replication sub-group.

3.2.6 Replication analysis using QUASAR2

Because rs13260246 reached genome-wide significance for vomiting in patients treated with XELOX, I also sought replication for this biomarker using data from 930 patients enrolled in QUASAR2 (Chapter 2, Section 2.2.2). Three patients had missing data and were excluded, leaving 927 to be analysed. The imputation score for rs13260246 was 0.96. Vomiting was graded using the CTCAE scale and patients with grades 2-5 (22%) were compared to those with grades 0-1.

3.2.7 Bioinformatic analyses

The GTEx project database was used to identify QTLs for relevant SNPs (Chapter 2, Section 2.5.4.2). Significance for tissue association was set at $P<1.0\text{x}10^{-3}$ (i.e. Bonferroni correction for 49 tissues [0.05/49]). Fine mapping was used for SNPs at significant loci using PAINTOR (Kichaev *et al*, 2014). Credible sets of causal SNPs were assembled for 95% coverage. Regional association plots were generated using the online version of Locuszoom (Chapter 2, Section 2.4.1).

3.3 Results

3.3.1 Rates of toxicity

There were significant differences in the incidences of toxicities associated with different chemotherapy regimens and cetuximab administration in COIN and COIN-B (**Table 3.1**). Notably, patients treated with FOLFOX had a significantly higher incidence of neutropenic sepsis, neutropenia and stomatitis, those with XELOX had a higher incidence of nausea, and those with cetuximab had a higher incidence of skin rash, HFS and diarrhoea (Chapter 2, **Table 2.2**). In view of this, patients were analysed for associations with genetic biomarkers after segregation by chemotherapy treatment and cetuximab status (**Figure 3.1**). In total, 385 patients with genotyping data were treated with FOLFOX, 360 with FOLFOX + cetuximab, 707 with XELOX and 348 with XELOX + cetuximab.

3.3.2 Genomic inflation and power considerations

4 million SNPs were analysed for a relationship with any toxicity and ten individual toxicities in each of the four patient groups. QQ plots of observed versus expected χ^2-test statistics showed no evidence of inflation of test statistics for all 40 GWAS performed (lambda range 0.99-1.02). There was 70% power to detect a mean OR of 4.3 (range 3-6) for any toxicity and 5.9 (2-39) for individual toxicities (**Table 3.2**). GWAS 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 were not performed due to having insufficient power.

Table 3.1 Significance of cetuximab and chemotherapy regimen on toxicities in patients from COIN and COIN-B

	Cetuximab	Chemotherapy regimen	Interaction
Any Toxicity	$1.5x10^{-5}$	0.99	$1.7x10^{-2}$
Individual toxicities			
Diarrhoea	$1.8x10^{-3}$	0.27	0.64
Neutropenic sepsis	$2.2x10^{-6}$	$3.3x10^{-7}$	$8.5x10^{-5}$
Peripheral neuropathy	0.24	$3.7x10^{-2}$	0.98
HFS	$7.2x10^{-10}$	$5.2x10^{-3}$	0.12
Neutropenia	$3.3x10^{-3}$	$2.0x10^{-16}$	$1.2x10^{-3}$
Lethargy	0.72	0.36	0.08
Stomatitis	$3.9x10^{-12}$	$5.4x10^{-5}$	$7.8x10^{-5}$
Nausea	0.38	$5.9x10^{-5}$	0.41
Vomiting	0.18	$2.3x10^{-3}$	0.08
Rash	$2.0x10^{-16}$	0.90	$3.0x10^{-2}$

HFS - Hand-foot syndrome.

Table 3.2 Detectable odds ratios at 70% power

	FOLFOX n=385	FOLFOX + cetuximab n=360	XELOX n=707	XELOX + cetuximab n=348
Any Toxicity	3.8	6.4	2.5	4.4
Individual toxicities				
Diarrhoea	3.8	3.5	2.6	3.5
Neutropenic sepsis	7.4	5.6	NA	NA
Peripheral neuropathy	5.2	7.5	2.9	5.3
HFS	38.5	4.6	4.0	4.6
Neutropenia	3.5	3.4	5.2	NA
Lethargy	3.3	3.4	2.4	3.6
Stomatitis	4.8	3.6	5.6	7.1
Nausea	5.3	5.0	2.7	4.2
Vomiting	7.9	6.2	3.2	6.1
Rash	NA	3.7	19.4	3.6

Detectable odds ratios are shown for a univariate additive model, based on a minor allele frequency of 20% and a significance of P=5.0x10^{-8}. NA – group sizes were too small to be calculated. HFS - Hand-foot syndrome.

3.3.3 Relationship between SNP genotype and any toxicity

No SNPs were associated with any toxicity at genome-wide significant levels (P<5.0x10^{-8}). SNPs at 27 loci were associated at suggestive levels (P<1.0x10^{-5}) (5 with FOLFOX, 8 with FOLFOX + cetuximab, 7 with XELOX and 7 with XELOX + cetuximab) (**Figure 3.2**); however, no lead SNPs were independently replicated 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 (**Table 3.3**).

3.3.4 Relationship between SNP genotype and individual toxicities

3.3.4.1 Vomiting

rs13260246 at 8q21.3 was associated with vomiting in patients treated with XELOX (OR=5.0, 95% CI=3.0-8.3, P=9.8x10^{-10}) (**Figure 3.3**). However, the association was not replicated in COIN and COIN-B patients treated with XELOX + cetuximab (P=0.72), nor in those receiving FOLFOX (P=0.35), with >90% power (**Table 3.4**). I also failed to replicate the association for rs13260246 with vomiting in the QUASAR2 trial of capecitabine alone versus capecitabine + bevacizumab for stage II and III CRC, regardless of treatment arm studied (with >99% power) (**Table 3.4**). rs13260246 was an eQTL for *SLC26A7* and five other genes **(Figure 3.3)**. SNPs at 15 loci had suggestive associations with vomiting but none were independently replicated.

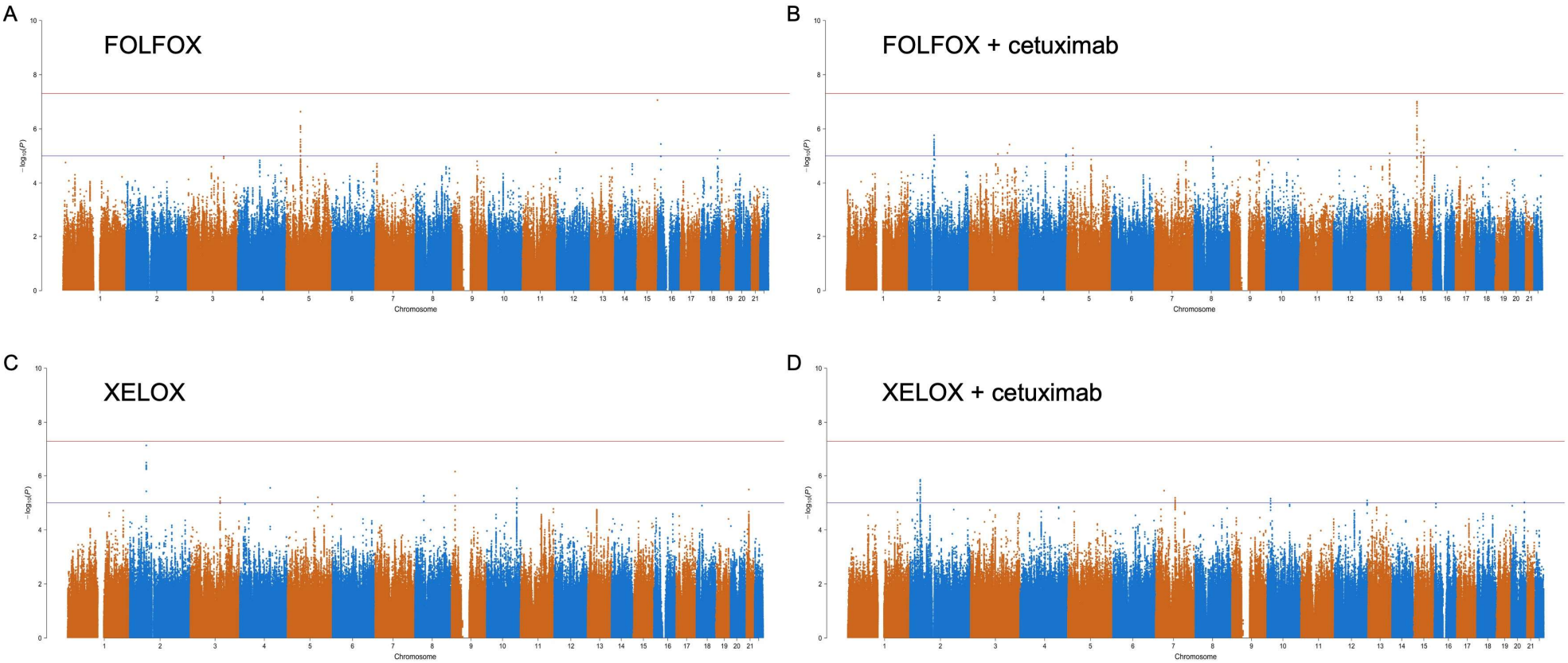

Figure 3.2 Manhattan plots of the relationship between SNP genotype and any toxicity. Patients treated with **(A)** FOLFOX (n=385), **(B)** FOLFOX + cetuximab (n=360), **(C)** XELOX (n=707) and **(D)** XELOX + cetuximab (n=348). The red line indicates a genome-wide significance threshold of P=5.0x10^{-8} and the blue line indicates a suggestive significance threshold of P=1.0x10^{-5}.

Table 3.3 SNPs with suggestive associations for any toxicity and independent replications

			GWAS			Replication chemo	Replication cetuximab status
	Lead SNP	Cytoband	OR	95% CI	*P*-value	*P*-value	*P*-value
FOLFOX	rs7181923	15q26.3	0.4	0.3-0.6	$8.6x10^{-8}$	0.58	0.19
	rs34265761	5q11.2	0.4	0.3-0.6	$2.3x10^{-7}$	0.16	0.39
	rs153081	16p13.12	0.5	0.4-0.7	$3.6x10^{-6}$	0.65	NA
	rs8090986	18q23	0.5	0.4-0.7	$6.1x10^{-6}$	NA	0.35
	rs12276840	11q25	2.0	1.5-2.7	$7.5x10^{-6}$	NA	NA
FOLFOX + cetuximab	rs35157945	15q14	0.2	0.1-0.4	$9.8x10^{-8}$	0.48	NA
	rs76301897	2q11.2	0.3	0.2-0.5	$1.7x10^{-6}$	0.41	0.55
	rs73015484	3q26.1	0.2	0.1-0.4	$3.8x10^{-6}$	0.20	0.78
	rs74946974	8q13.2	0.2	0.1-0.4	$4.6x10^{-6}$	0.53	0.51
	rs58842897	5p14.1	0.2	0.1-0.4	$5.2x10^{-6}$	0.40	NA
	rs75434917	20p11.22	0.2	0.1-0.4	$5.9x10^{-6}$	NA	0.55
	rs4503663	13q33.3	2.3	1.6-3.3	$8.0x10^{-6}$	0.16	NA
	rs6827299	4q35.2	0.4	0.3-0.6	$8.9x10^{-6}$	NA	0.06
XELOX	rs17709614	9p22	0.6	0.5-0.7	$7.0x10^{-7}$	NA	0.44
	rs72621832	4q27	0.4	0.3-0.6	$2.8x10^{-6}$	NA	0.22
	rs1932542	10q26.1	1.9	1.5-2.5	$2.9x10^{-6}$	NA	0.18
	rs34330891	21q21	0.6	0.5-0.7	$3.2x10^{-6}$	NA	0.81
	rs11786456	8p12	1.8	1.4-2.3	$5.5x10^{-6}$	NA	NA
	rs9812615	3q13.33	2.0	1.5-2.7	$6.5x10^{-6}$	NA	NA
	rs150312337	2p14	0.4	0.3-0.6	$7.6x10^{-6}$	NA	0.61

XELOX + cetuximab	rs4438529	2p22.1	0.4	0.3-0.6	$1.4x10^{-6}$	0.18	NA
	rs315866	7p14.3	0.2	0.1-0.4	$3.6x10^{-6}$	0.22	NA
	rs6716820	2p23.2	0.3	0.2-0.5	$4.4x10^{-6}$	0.41	1.0
	rs71897151	10p13	0.4	0.3-0.6	$7.0x10^{-6}$	NA	1.0
	rs6966363	7q21.11	0.4	0.3-0.6	$7.9x10^{-6}$	NA	1.0
	rs61942090	12q24.33	2.3	1.6-3.3	$8.1x10^{-6}$	NA	NA
	rs35775456	20q13.2	0.2	0.1-0.4	$9.7x10^{-6}$	NA	0.06

Replication chemo - Replication in the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status. Replication cetuximab status - Replication in the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status. NA = Replication odds ratio in opposite direction to the GWAS odds ratio, OR = Odds ratio, CI = Confidence intervals.

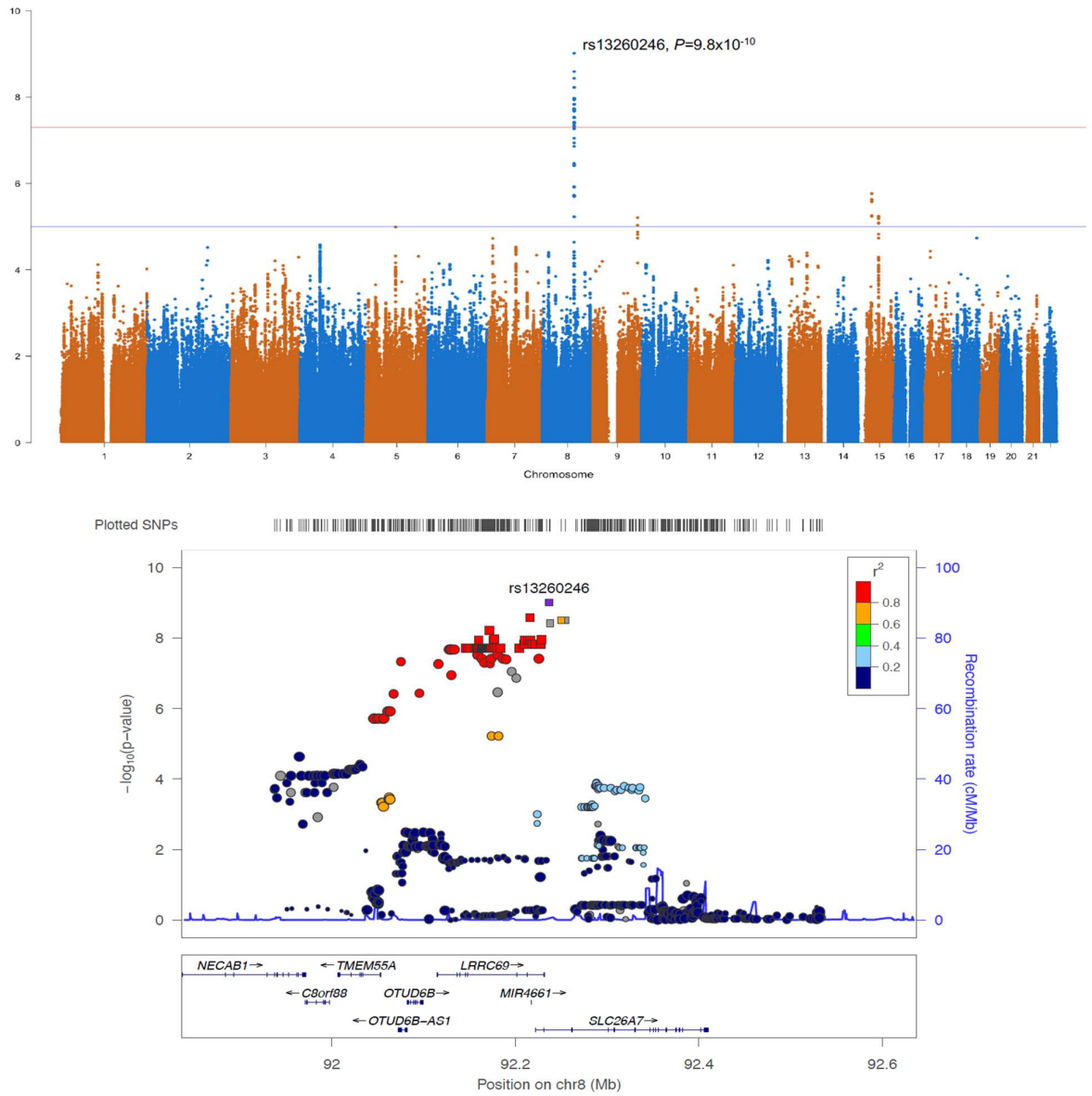

Figure 3.3 Regional association plots of the relationship between SNP genotype and vomiting in patients treated with XELOX. (A) Manhattan plot for the vomiting GWAS. The red line corresponds to a P=5.0x10^{-8} and the blue line P=1.0x10^{-5}. **(B)** Regional plot for the 8q21.3 association with vomiting. Plot shows results of the analysis for single-nucleotide polymorphisms (SNPs) and recombination rates. −log10(P) (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The sentinel SNP (purple) is labelled by its rsID. The colour intensity of each symbol reflects the extent of linkage disequilibrium (LD) with the sentinel SNP, deep blue (r2=0) through to dark red (r2=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 37 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. Fine-mapping identified a credible set of 70 SNPs with rs13260246 having the highest posterior probability of 0.14.

Table 3.4 rs13260246 associated with vomiting in patients treated with XELOX and analyses of replication cohorts

Trial	Treatment	Total patients	Patients G0-1 for vomiting			Patients G2-5 for vomiting			OR	95% CI	*P*-value
			wild type	heterozygous	homozygous	wild type	heterozygous	homozygous			
COIN and COIN-B	XELOX	695	555	54	0	58	26	2	5.0	3.0-8.3	$9.8x10^{-10}$
	XELOX + cetuximab	341	269	37	0	31	3	1	1.2	0.5-3.2	0.72
	FOLFOX	378	318	35	1	23	1	0	0.4	0.1-2.9	0.35
QUASAR2	Capecitabine	440	315	40	0	82	3	0	0.3	0.1-1.0	$2.7x10^{-2}$*
	Capecitabine + bevacizumab	487	322	48	1	96	18	2	1.4	0.8-2.4	0.22

Reference allele = C, OR = Odds ratio, CI = Odds ratio confidence intervals. *OR in opposite direction for replication. Total patients excludes those with missing genotypes (12 treated with XELOX, 7 treated with XELOX + cetuximab and 7 treated with FOLFOX). rs13260246 has a minor allele frequency of 0.07 in gnomAD v.2.1.1 (Europeans non-Finnish).

3.3.4.2 Diarrhoea

SNPs at 21 loci had suggestive associations with diarrhoea (**Figure 3.4**); however, only rs6030266 at 20q13.12 in patients treated with XELOX + cetuximab (OR=0.4, 95% CI=0.28-0.58, P=5.7x10^{-7}) was replicated in patients receiving FOLFOX + cetuximab (OR=0.7, 95% CI=0.5-0.9, P=3.6x10^{-2}); Pooled P=3.2x10^{-7} (**Table 3.5**). rs6030266 maps to intron 8 of the gene encoding protein tyrosine phosphatase receptor type T (*PTPRT*) (**Figure 3.5A**).

3.3.4.3 HFS

SNPs at 13 loci had suggestive associations with HFS (**Figure 3.4**). Only rs1546161 at 1q21.2 in patients treated with FOLFOX (OR=17.8, 95% CI=5.1-62.0, P=5.9x10^{-6}) was replicated in those receiving XELOX (OR=1.7, 95% CI=1.1-2.7, P=2.5x10^{-2}); Pooled P=2.5x10^{-6} (**Table 3.5**). rs1546161 maps to B-Cell Lymphoma 9 (*BCL9*) and was an eQTL for *GJA5* (**Figure 3.5B**).

3.3.4.4 Neutropenia

SNPs at 13 loci had suggestive associations with neutropenia (**Figure 3.4**). Only rs9601722 at 13q31.1 in patients treated with FOLFOX + cetuximab (OR=3.4, 95% CI=2.0-5.7, P=5.2x10^{-6}) was replicated in those receiving FOLFOX (OR=1.7, 95% CI=1.1-2.9, P=3.6x10^{-2}); Pooled P=3.0x10^{-6} (**Table 3.5**). rs9601722 maps to a lncRNA (*LOC105370284*).

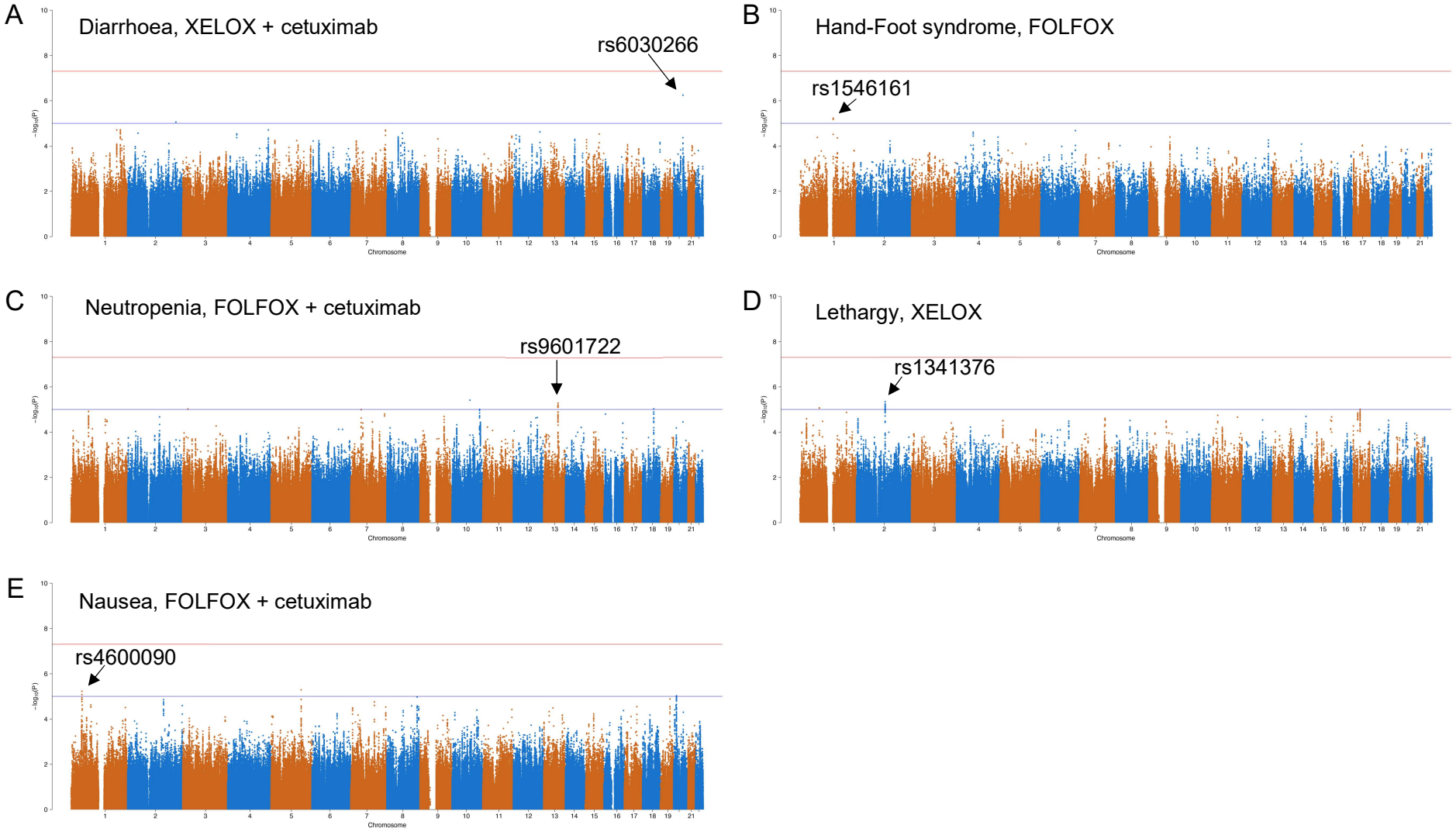

Figure 3.4 Manhattan plots of the relationship between SNP genotype and (A) Diarrhoea in patients treated with XELOX + cetuximab, (B) Hand-foot syndrome (HFS) in patients treated with FOLFOX, (C) Neutropenia in patients treated with

FOLFOX + cetuximab, (D) Lethargy in patients treated with XELOX, and, (E) Nausea in patients treated with FOLFOX + cetuximab. Replicated SNPs were: (**A)** rs6030266 at 20q13.12 (P=5.7x10^{-7}) which replicated in patients treated with FOLFOX + cetuximab (P=3.6x10^{-2}); Pooled P=3.2x10^{-7}. (**B)** rs1546161 at 1q21.2 (P=5.9x10^{-6}) which replicated in patients treated with XELOX (P=2.5x10^{-2}); Pooled P=2.5x10^{-6}. (**C)** rs9601722 at 13q31.1 (P=5.2x10^{-6}) which replicated in patients treated with FOLFOX (P=3.6x10^{-2}); Pooled P=3.0x10^{-6}. (**D)** rs13413764 at 2q14.3 (P=4.5x10^{-6}) which replicated in patients treated with FOLFOX (P=9.2x10^{-3}); Pooled P=7.5x10^{-7}. (**E)** rs4600090 at 1p33 (P=5.9x10^{-6}) which replicated in patients treated with FOLFOX (P=4.2x10^{-2}); Pooled P=4.0x10^{-6}. The red line indicates a genome-wide significance threshold of P=5.0x10^{-8} and the blue line indicates suggestive significance threshold at P=1.0x10^{-5}

Table 3.5 Replicated SNPs associated with individual toxicities

Toxicity	Treatment group	Lead SNP	Cytoband	GWAS			Replication chemo	Replication cetuximab status	Combined
				OR	95% CI	*P*-Value	*P*-value	*P*-value	*P*-value
Diarrhoea	XELOX + cetuximab	rs6030266	20q13.12	0.4	0.3-0.6	$5.7x10^{-7}$	0.33	$3.6x10^{-2}$	$3.2x10^{-7}$
HFS	FOLFOX	rs1546161	1q21.2	17.8	5.1-62	$5.9x10^{-6}$	0.13	$2.5x10^{-2}$	$2.5x10^{-6}$
Neutropenia	FOLFOX + cetuximab	rs9601722	13q31.1	3.4	2.0-5.7	$5.2x10^{-6}$	$3.6x10^{-2}$	NA	$3.0x10^{-6}$
Lethargy	XELOX	rs13413764	2q14.3	1.8	1.4-2.3	$4.5x10^{-6}$	NA	$9.2x10^{-3}$	$7.5x10^{-7}$
Nausea	FOLFOX + cetuximab	rs4600090	1p33	4.0	2.2-7.2	$5.9x10^{-6}$	$4.2x10^{-2}$	0.55	$4.0x10^{-6}$

Replication chemo - Replication in the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status. Replication cetuximab status - Replication in the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status. Combined *P* - Pooled *P*-value of GWAS and replicated cohorts (excludes cohort which was not replicated). OR = Odds ratio, CI = confidence intervals, NA = OR in the opposite direction to the GWAS. HFS- Hand-foot syndrome.

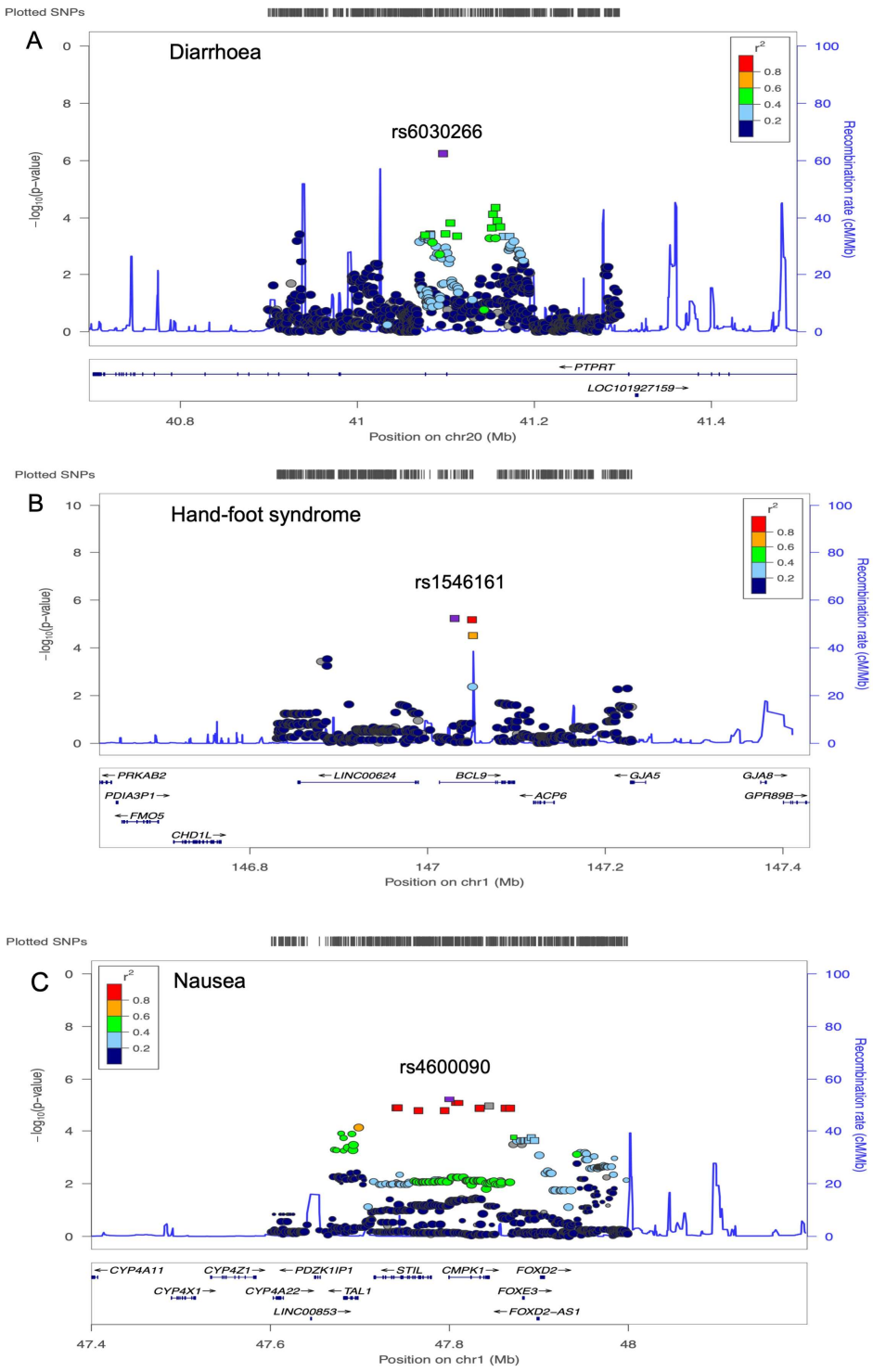

Figure 3.5 Regional plots of (A) the 20q11.2 association with diarrhoea, (B) the 1q21.2 association with Hand-foot syndrome (HFS) and (C) the 1p33 association with nausea. Plots show results of the analysis for single-nucleotide

polymorphisms (SNPs) and recombination rates. $-\log_{10}(P)$ (y axis) of the SNPs are shown according to their chromosomal positions (x axis). 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 37 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. Fine-mapping identified a credible set of **(A)** 20 SNPs with rs6030266 having the highest posterior probability of 0.87, **(B)** 3 SNPs with rs1546161 having the highest posterior probability of 0.47 and **(C)** 19 SNPs with rs4600090 having the highest posterior probability of 0.14 (SNPs that form the 95% credible set are denoted by squares).

3.3.4.5 Lethargy

SNPs at 12 loci had suggestive associations with lethargy (**Figure 3.4**); however, only rs13413764 at 2q14.3 in patients treated with XELOX (OR=1.8, 95% CI=1.4-2.3 P=4.5x10^{-6}) was replicated in those receiving FOLFOX (OR=1.5, 95% CI=1.1-2.1, P=9.2x10^{-3}); Pooled P=7.5x10^{-7} (**Table 3.5**). rs13413764 maps to an intergenic region.

3.3.4.6 Nausea

SNPs at 12 loci had suggestive associations with nausea (**Figure 3.4**). However, only rs4600090 at 1p33 in patients treated with FOLFOX + cetuximab (OR=4.0, 95% CI=2.2-7.2, P=5.9x10^{-6}) was replicated in those receiving FOLFOX (OR=2.0, 95% CI=1.1-4.0, P=4.2x10^{-2}); Pooled P=4.0x10^{-6} (**Table 3.5**). rs4600090 was an eQTL for *CMPK1*, *FOXE3* and *PDZK1IP1* (**Figure 3.5C**).

3.3.4.7 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 replicated.

3.3.4.8 Association between genes and neutropenia

Four genes were significantly associated with neutropenia (using a Bonferroni corrected threshold of P<2.5x10^{-6}). Of these, Maestro Heat-Like Repeat Family Member 5 (*MROH5*), found in patients treated with XELOX (P=6.6x10^{-7}), was replicated in those receiving XELOX + cetuximab (P=3.3x10^{-2}); Pooled P=3.7x10^{-7} (**Table 3.6**). 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=1.0x10^{-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=1.4x10^{-6}) as compared to those receiving XELOX + cetuximab (lead SNP rs12056882 OR=4.4, 95% CI=1.4-14, P=1.0x10^{-2}). After additional imputation was performed (to increase the SNP density), rs67817056 became the most significantly associated SNP in patients treated with XELOX + cetuximab (P=2.9x10^{-3}; **Figure 3.6**). Neither rs76380775 nor rs12056882 were associated with neutropenic sepsis or white blood cell count. rs12056882 was an sQTL for *PTP4A3* (which lies 1.37kb downstream of *MROH5*).

3.3.5 Genes and gene sets associated with other toxicities

Of the other significant genes, three (all mapping to 8q21.3) were associated with vomiting and one was associated with stomatitis (**Table 3.6**). For gene sets 2, 8 and 3 gene sets were associated with any toxicity, lethargy and vomiting, respectively. However, all these genes and gene sets failed independent replication (**Tables 3.7 and 3.8**).

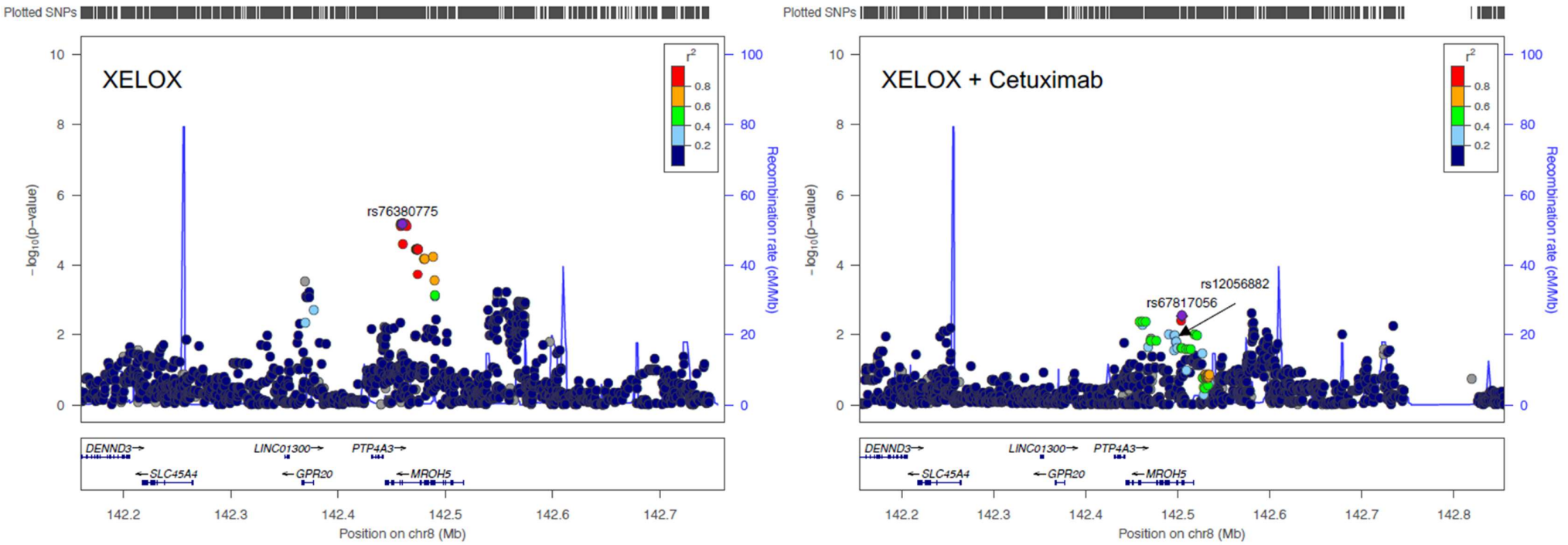

Figure 3.6 MROH5 regional plots associated with neutropenia. Plots show results of the analysis for single-nucleotide polymorphisms (SNPs) and recombination rates. $-\log_{10}(P)$ (y axes) of the SNPs are shown according to their chromosomal positions (x axes). The sentinel SNP (purple) in each analysis 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 37 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. The association with *MROH5* appeared to be due to independent sets of SNPs: rs76380775 was the lead SNP in patients treated with XELOX (P=1.4x10^{-6}), and rs12056882 was lead SNP in patients treated with XELOX + cetuximab (P=1.0x10^{-2}). After additional imputation (to increase the SNP density), rs67817056 became the most significantly associated SNP in patients treated with XELOX + cetuximab (P=2.9x10^{-3}).

Table 3.6 MAGMA gene analyses for individual toxicities

Toxicity	Treatment group	Gene	*P*-value	Replication chemo *P*-value	Replication cetuximab status *P*-value	Pooled *P*-value
Neutropenia	FOLFOX	*RPL17-C18orf32*	8.9×10^{-7}	0.57	0.53	-
		C18orf32	1.3×10^{-6}	0.56	0.51	-
		RPL17	1.5×10^{-6}	0.56	0.52	-
	XELOX	***MROH5***	**6.6×10^{-7}**	**3.3×10^{-2}**	**0.09**	**3.7×10^{-7}**
Stomatitis	FOLFOX	*SCAF4*	1.3×10^{-6}	0.07	0.61	-
Vomiting	XELOX	*LRRC69*	1.2×10^{-7}	0.77	0.73	-
		SLC26A7	4.3×10^{-7}	0.81	0.60	-
		PIP4P2	9.7×10^{-7}	0.94	0.34	-

Replication chemo - Replication in the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status. Replication cetuximab status - Replication in the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status. 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 replicated 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).

Table 3.7 MAGMA gene set analyses for any toxicity

Treatment	Pathway	GWAS *P*-value	Replication chemo *P*-value	Replication cetuximab status *P*-value
XELOX + cetuximab	Negative regulation of cellular response to insulin stimulus	$1.7x10^{-6}$	0.07	0.71
	Negative regulation of insulin receptor signalling pathway	$1.8x10^{-6}$	0.09	0.77

Significance was a Bonferroni corrected threshold of $P<5.6x10^{-6}$. Replication chemo - Replication in the COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status. Replication cetuximab status - Replication in the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status. No results passed replication at $P<0.05$.

Table 3.8 MAGMA gene set analyses for individual toxicities

Toxicity	Treatment	Pathway	GWAS *P*-value	Replication chemo *P*-value	Replication cetuximab status *P*-value
Lethargy	FOLFOX + cetuximab	Dopamine transport	$4.3x10^{-6}$	0.88	0.77
	XELOX	Purine nucleoside biosynthetic process	$1.2x10^{-7}$	0.30	0.58
		Purine ribonucleoside biosynthetic process	$1.2x10^{-7}$	0.30	0.58
		Actin-dependent ATPase activity	$7.6x10^{-7}$	0.35	0.68
		Ral GTPase binding	$3.6x10^{-6}$	0.46	0.80
		G protein-coupled receptor internalization	$3.8x10^{-6}$	0.34	0.50
		Nucleoside biosynthetic process	$4.2x10^{-6}$	0.48	0.51
	XELOX + cetuximab	Endoplasmic reticulum calcium ion homeostasis	$4.6x10^{-7}$	0.19	0.79
Vomiting	FOLFOX	Mitogen-activated protein kinase kinase binding	$2.3x10^{-6}$	0.99	0.90
	FOLFOX + cetuximab	Nuclear membrane	$3.7x10^{-6}$	0.25	0.64
	XELOX	Euchromatin	$4.8x10^{-6}$	0.25	0.21

Genome-wide significance was a Bonferroni corrected significance threshold of $P<5.6x10^{-6}$. Replication chemo – Replication in the

COIN and COIN-B group with the same chemotherapy regimen but alternative cetuximab status. Replication cetuximab status - Replication in the COIN and COIN-B group with the alternative chemotherapy regimen but with the same cetuximab status. No results replicated at $P<0.05$.

3.3.6 Lack of confounding effect for rare *DPYD* variants

It has previously been shown that two rare variants in *DPYD* (Asp949Val and IVS14+1G>A) were associated with a range of toxicities in COIN and COIN-B (Madi *et al*, 2018). Of the 1,800 patients in the GWAS, 22 carried Asp949Val and 17 carried IVS14+1G>A. Excluding these patients made no significant differences to the strengths of associations reported *herein* (**Table 3.9**).

3.3.7 Evaluation of previously purported associations

A previous GWAS for toxicity to 5FU or FOLFOX in patients with CRC identified two SNPs associated with mucositis, two with diarrhoea and three with haematological toxicities (Fernández-Rozadilla *et al*, 2013). All these SNPs failed to replicate in COIN and COIN-B (**Table 3.10**), despite having adequate power.

Table 3.9 Lack of confounding effect of rare toxicity-associated *DPYD* variants on biomarkers identified *herein*

Lead SNP	Toxicity	Treatment group	Number of patients excluded	*P*-value	
				Before exclusion	After exclusion
rs13260246	Vomiting	XELOX	13	9.8×10^{-10}	2.2×10^{-9}
rs6030266	Diarrhoea	XELOX + cetuximab	6	5.7×10^{-7}	1.1×10^{-6}
rs1546161	HFS	FOLFOX	15	5.9×10^{-6}	7.0×10^{-6}
rs9601722	Neutropenia	FOLFOX + cetuximab	5	5.2×10^{-6}	7.0×10^{-6}
rs13413764	Lethargy	XELOX	13	4.5×10^{-6}	1.6×10^{-5}
rs4600090	Nausea	FOLFOX + cetuximab	5	5.9×10^{-6}	7.4×10^{-6}
rs76380775 (*MROH5*)	Neutropenia	XELOX	13	1.4×10^{-6}	1.6×10^{-6}

P-values before and after excluding patients with the rare *DPYD* variants Asp949Val and IVS14+1G>A. HFS- Hand-foot syndrome.

Table 3.10 Lack of replication of loci identified by Fernández-Rozadilla *et al.* (2013)

Toxicity	SNP	Fernández-Rozadilla *et al.* (2013) Pooled *P*-value	COIN and COIN-B			
			FOLFOX *P*-value	FOLFOX + cetuximab *P*-value	XELOX *P*-value	XELOX + cetuximab *P*-value
Diarrhoea						
	rs10876844	1.0×10^{-2}	0.87	0.72	0.57	0.66
	rs10784749	1.7×10^{-2}	0.69	0.70	0.08	0.97
Haematological						
	rs7325568	2.3×10^{-4}	0.89	0.38	0.49	0.45
	rs4243761	2.8×10^{-3}	0.96	0.97	NA	0.20
	rs17626122	4.2×10^{-2}	0.74	0.13	0.31	0.42
Mucositis						
	rs2465403	9.4×10^{-3}	0.65	0.52	0.52	0.31
	rs16857540	2.0×10^{-2}	0.42	0.44	0.40	0.96

All loci failed to replicate in COIN and COIN-B. COIN did not record mucositis (inflammation of the mouth and digestive tract) so SNPs were replicated using stomatitis (inflammation of the mouth). rs10876844 was not genotyped in COIN so a proxy SNP, rs2555036, was used (r^2=0.69). The power to replicate the stomatitis loci was >99%, the diarrhoea loci >99% and the haematological loci 40% in those treated with FOLFOX and 7% in those treated with XELOX. Power was calculated using the SNP minor allele frequency in COIN and COIN-B, and the odds ratio from Fernández-Rozadilla *et al.* (2013). NA = odds ratio in opposite direction.

3.4 Discussion

3.4.1 Exploring the mechanism underlying the association of *MROH5* with neutropenia

MROH5 was identified from MAGMA gene analyses as associated with neutropenia at genome-wide significant levels in patients treated with XELOX and was independently replicated 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 an 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, which is not in LD with rs12056882. *PTP4A3* represents a strong causal candidate for neutropenia as treatment of mice with a *PTP4A3* derived peptide reduced endotoxemia induced septic shock (Tang *et al*, 2009). *PTP4A3* expression has also been associated with poor prognosis in CRC possibly due to a role in metastasis and tumour invasion (Zimmerman *et al*, 2013; Saha *et al*, 2001), and has been implicated in resistance to chemotherapy (Csoboz *et al*, 2018; Hollander *et al*, 2016).

3.4.2 Evaluating the association between rs13260246 and vomiting

There was also a clear signal for rs13260246 associated with vomiting in patients treated with XELOX. However, this association was not replicated in patients treated with XELOX + cetuximab, nor in those receiving FOLFOX. While rs13260246 was significant in patients treated with capecitabine +/- bevacizumab from the QUASAR2 trial, the odds ratio was in the opposite direction. It has been proposed that loci effects can flip-flop particularly in GWAS as they miss some genetic complexity (Lin

et al, 2007). However, without any further mechanistic evidence to support this, my data suggest that rs13260246 could be a false-positive or 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 (Kim *et al*, 2005), 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 (Fujiwara *et al*, 2013; Andreyev *et al*, 2020; Hartmann, 2015). *SLC26A7* is also expressed in parietal cells and genetic deletion results in decreased gastric acid secretion (Xu *et al*, 2009; Petrovic *et al*, 2003). Both thyroid and gastric dysfunction can cause vomiting (Sweet *et al*, 2010; Raufman *et al*, 1983). Therefore, *SLC26A7* represents a potential biological candidate for vomiting but lacks genetic replication at present.

3.4.3 Suggestive significance loci

In total, SNPs at 139 loci had evidence for suggestive associations for any toxicity or individual toxicities and lead SNPs at five of these were replicated at nominally significant levels. However, if a more stringent correction for 139 replication tests was applied, none of the five would have passed the adjusted significance threshold. Further replication 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 8 of *PTPRT*, a tumour suppressor gene that functions as part of the JAK/STAT pathway (Hsu *et al*, 2018). rs1546161 was associated with HFS and maps to *BCL9*, overexpression of which has been linked to disrupted *wnt* signalling (Takada *et al*, 2012). 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 5FU phosphorylation and linked to 5FU sensitivity (Yasuno *et al*, 2013). 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 (Fredrikson *et al*, 1994). rs9601722 associated with neutropenia and rs13413764 with lethargy did not lie within protein coding gene regions.

3.4.4 Evaluation of other significant genes and gene sets

As all other genes and gene sets failed to replicate across groups their likelihood of being true associations is reduced. However, as the replication groups had non-identical therapies, there is a chance that the genes and gene sets are highly therapy specific. It is noteworthy that from a biological standpoint, none of the genes or gene sets have support in the literature for association with their respective toxicity, suggesting they are likely to be false positives.

3.4.5 Study limitations

One design decision was to analyse by treatment subgroup rather than by analysing the whole cohort and adjusting for treatment effects through covariates. I choose this method to identify treatment specific markers which may otherwise be missed if the cohort was analysed all together. However, this did limit the power available for individual GWAS, especially for variants with low odds ratios (<2). Furthermore, my attempts to replicate any findings was limited by groups with similar, but non-identical, therapies.

3.4.6 Conclusions and follow-on studies

After conducting 40 GWAS on large patient cohorts with well characterised clinical data, I 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, I failed to replicate loci previously suggested to be associated with toxicity to FOLFOX identified from another GWAS (Fernández-Rozadilla, 2013). Meta-analyses can be performed to improve the power to detect associations with lower effect sizes and I have carried out such analyses in Chapters 4 and 5.

Further analyses of *MROH5* and *PTP4A3* with neutropenia are warranted to establish the mechanism of effect. Identification of a XELOX only cohort to replicate the vomiting loci would also be ideal, to determine if the locus is treatment specific or a false positive.

4 Meta-analyses of COIN and COIN-B to identify loci associated with toxicity to FOLFOX and XELOX chemotherapy regimens

4.1 Introduction

4.1.1 Meta-analyses

While GWAS have increased our understanding of complex inherited traits, often loci will fail to validate in replication studies (Ioannidis, 2007; Pearson and Manolio, 2008). An approach to alleviate this is meta-analysis, a statistical method used to integrate the results of several studies into a single conclusion (Glass, 1976). GWAS meta-analyses benefit from increased statistical power and precision (Trikalinos *et al*, 2008), being more likely to identify true loci because their effects must already have been present in several of the cohorts meta-analysed (Panagiotou *et al*, 2013).

In a meta-analysis, it is important to consider heterogeneity between studies to ensure an appropriate model is used (Egger *et al,* 1997). There are numerous statistics that can be used but all work on the same principle (Zeggini and Ioannidis, 2009). A low heterogeneity score indicates there is minimal inter-study variation, so a fixed effects model will provide the most power (Nakaoka and Inoue, 2009). A high heterogeneity score indicates the studies are heterogeneous which must be accounted for during analysis. The most common method for this is to use a random effects model.

4.1.1.1 Fixed effects model

Fixed effect models assume that for each tested variable, there is one true effect. Associations are expected to be consistent across studies, only differing due to intra-

study error (Hunter and Schmidt, 2000). During the meta-analysis, individual studies are weighted depending on their sample size, with results from larger cohorts being weighted more highly. This increases precision, as false positives are more prevalent in smaller cohorts, which are allowed to bias the overall results less (Mikolajewicz and Komarova, 2019).

4.1.1.2 Random effects model

Random effect models can account for heterogeneity between studies. The test assumption is that the true effect of a variable can vary (DerSimonian and Laird, 1986). This allows for underlying differences in the studies to be accounted for such as treatment or cancer stage. Individual studies are not weighted in the meta-analysis and results are instead influenced by the number of studies the effect is present in and the association strength in each (Mikolajewicz and Komarova, 2019). Random effect models work well when meta-analysing clinical studies since patient inclusion criteria can differ significantly (Borenstein *et al*, 2010).

4.1.2 HFS

Toxicity from chemotherapy may result in treatment discontinuation or dose reduction affecting the prospect of a cure in patients with cancer. Patients treated with XELOX often develop HFS, in which small amounts of the chemotherapeutic agent leaks out of capillaries into the hands and feet and damages the surrounding tissues (Milano *et al*, 2008). HFS is characterised by erythema, blisters, peeling of the skin on the hands and feet and at higher grades pain that limits daily living activities. HFS has been suggested to be a biomarker of treatment efficacy with *post-hoc* analyses from clinical trials of colorectal and breast cancer patients finding that

grade 1+ HFS was associated with improved overall and progression-free survival (Stintzing *et al,* 2011; Zielinski *et al*, 2016). Established risk factors for HFS include being older, female, having pre-existing peripheral neuropathy, circulation problems and diabetes (Diasio, 2000; Kooner *et al*, 2011). Common genetic variants in *TYMS* and *MTHFR* have been associated with increased risk of HFS toxicity, albeit with low effect sizes (Rosmarin *et al*, 2015; Lin *et al*, 2019).

4.1.3 ST6 β-galactoside α-2,6-sialyltransferase 1 (*ST6GAL1*)

Genetic variation in *ST6GAL1* is associated with a risk of developing type-2 diabetes (T2D) (Mahajan *et al*, 2018; Kaburagi, 2017). *ST6GAL1* catalyses the addition of α2,6-linked sialic acids onto key surface glycoproteins. Increases in α2,6-linked sialic acids have been linked to inflammatory conditions (Yasukawa *et al*, 2005) and *ST6GAL1* deficiency leads to increased inflammatory cell production (Nasirikenari *et al*, 2006), granulocyte recruitment (Nasirikenari *et al*, 2010) and cytokine release (Nasirikenari *et al*, 2019). There is also substantial evidence that *ST6GAL1* plays an important role in cancer progression and it is overexpressed in numerous cancers including colorectal (Dorsett *et al*, 2021). High *ST6GAL1* expression has been associated with radioresistance and chemoresistance to several anticancer treatments, which ultimately leads to worse patient outcomes (Lee *et al*, 2008; Schultz *et al*, 2013, Britain *et al*, 2018; Duarte *et al*, 2021).

4.1.4 Aims

As in the previous chapter, I chose to analyse the cohort by treatment subgroup, in this chapter I aim to continue this methodology. I extended my analyses by meta-analysing those patients who received XELOX ± cetuximab and, separately,

FOLFOX ± cetuximab (Chapter 1, **Table 1.4**). After initial analyses, I then considered HFS as a biomarker of treatment efficacy. I also sought to confirm an association between *ST6GAL1* and T2D, and understand their inter-relationship with HFS by studying biomarkers of inflammation using data from the UK Biobank.

4.2 Materials and Methods

4.2.1 Patients and samples

As previously described (Chapter 2, Sections 2.2.1.4 and 2.2.1.5), toxicity data and SNP genotypes were available for 1,800 patients from COIN and COIN-B after QC. For analyses of response to treatment at 12 weeks, a further 210 patients had missing data and were excluded.

4.2.2 Toxicities assessed

Toxicities assessed were diarrhoea, neutropenic sepsis, peripheral neuropathy, HFS, neutropenia, lethargy, stomatitis, nausea, vomiting and rash. Patients with toxicities graded 2-5 (G2-5) were grouped and compared against those graded 0-1 (G0-1). Of note, for HFS, G3 is the maximum possible grade so, patients with G2-3 were compared against those with G0-1. A linear model was considered for the association between rs6783836 and HFS, to assess if rs6783836 was associated with toxicity severity.

4.2.3 Patient outcome

Assessment of response was also performed at 12 weeks. Response was defined as complete or partial response using RECIST 1.0 guidelines and no response was defined as stable or progressive disease. Overall survival (OS) was defined as the time from randomisation to death or date of last assessment.

4.2.4 GWAS

In Chapter 3, 4 million SNPs were analysed for a relationship with each toxicity under univariate models in patients that received XELOX (n=707), XELOX + cetuximab (n=348), FOLFOX (n=385) and FOLFOX + cetuximab (n=360). Here, I incorporated covariates associated at $P<0.05$ (**Table 4.1**) into the additive logistic models in Plink (v1.9) and meta-analysed those patients receiving XELOX ± cetuximab (n=1,055) and, separately, FOLFOX ± cetuximab (n=745). Meta-analyses were run under random effects models to account for heterogeneity caused by the effect of cetuximab on toxicity (**Chapter 3**, **Table 3.1**). Analyses were restricted to directly typed SNPs and imputed SNPs with imputation scores ≥0.8, a HWE $\geq 1.0 \times 10^{-6}$ and a MAF ≥0.05. Results were plotted in R studio using qqman and ggplot2. SNPs associated at genome-wide significance ($P<5.0 \times 10^{-8}$) were selected for further analyses.

4.2.5 Patient outcomes

For survival analyses, Cox proportional hazard regression models were used for both univariate and multivariate analyses. For analyses of response to treatment, logistic regression models were used for both univariate and multivariate analyses.

4.2.6 Gene and gene set analyses

MAGMA (de Leeuw *et al*, 2015) was used for gene and gene set analyses (Chapter 2 Section 2.5.3). Gene analyses were run under a snpwise univariate model imposing a Bonferroni corrected significance threshold of $P=2.5 \times 10^{-6}$. Gene set analyses were run under competitive models with a corrected significance threshold of $P=5.8 \times 10^{-6}$.

Table 4.1 Covariates used in the genome-wide association studies

	Sex	Creatinine clearance	Location of primary tumour	Age	Platelet count	WHO performance status
Diarrhoea	$3.3x10^{-2}$	$6.5x10^{-4}$	0.18	0.65	0.86	$3.8x10^{-2}$
Neutropenic sepsis	$4.6x10^{-4}$	0.53	0.07	0.15	0.12	0.07
Peripheral neuropathy	$3.1x10^{-2}$	0.71	0.53	0.70	$1.0x10^{-2}$	0.39
HFS	0.09	$4.6x10^{-2}$	0.46	$1.9x10^{-3}$	$1.8x10^{-2}$	0.82
Neutropenia	$2.5x10^{-4}$	$1.9x10^{-2}$	0.51	0.20	$2.9x10^{-3}$	0.13
Lethargy	$3.9x10^{-3}$	0.21	$3.6x10^{-2}$	0.18	0.74	$1.4x10^{-5}$
Stomatitis	$4.5x10^{-3}$	0.76	0.98	0.06	0.07	$3.0x10^{-3}$
Nausea	$5.3x10^{-4}$	$8.3x10^{-3}$	$8.5x10^{-3}$	$1.1x10^{-2}$	0.08	$3.0x10^{-2}$
Vomiting	0.09	$3.5x10^{-2}$	$2.6x10^{-2}$	$2.0x10^{-2}$	0.78	0.79
Rash	$1.3x10^{-3}$	0.50	0.09	$1.8x10^{-2}$	0.07	$1.1x10^{-2}$

Logistic model also had chemotherapy regimen and cetuximab as terms added. Covariates associated at $P<0.05$ were included in the genome-wide association studies. HFS- Hand-foot syndrome.

4.2.7 Power considerations

Power to detect toxicity effect sizes was calculated, based upon 70% power, P=5.0x10^{-8} and SNPs with MAFs=0.20. Under these conditions SNPs with a mean OR of 2.8 (range 2-4 dependent upon toxicity, **Table 4.2**) could be identified.

4.2.8 Independent replication

The association of rs6783836 with HFS was attempted to be replicated using data from 930 patients enrolled in QUASAR2 (Kerr *et al*, 2016). Three patients had missing data and were excluded, leaving 927 to be analysed. The imputation score for rs6783836 was 0.89. HFS was graded using the CTCAE scale and patients with G2-3 (46%) were compared to those with G0-1. Age was included as a covariate.

4.2.9 *ST6GAL1* and diabetes

Six hundred and fourteen SNPs spanning *ST6GAL1* were tested for an association with T2D in UK Biobank participants (17,384 cases and 317,887 controls as of 1 January 2021). The analysis was restricted to directly typed SNPs and imputed SNPs with imputation scores ≥0.8, a HWE ≥1.0×10^{-6} and a MAF ≥0.01. We also analysed the relationship between rs6783836 and diabetic skin lesions by logistic regression on 617 diabetic individuals with self-reported open sores and 6,605 diabetic controls (as of 1 July 2021).

Table 4.2 Patients with grade 2-5 CTCAE toxicities at 12 weeks and detectable odds ratios at 70% power

Toxicity	Frequency		Detectable odds ratio	
	XELOX ± cetuximab n (%)	FOLFOX ± cetuximab n (%)	XELOX ± cetuximab	FOLFOX ± cetuximab
Diarrhoea	288 (27)	187 (25)	2.1	2.5
Neutropenic sepsis	6 (1)	63 (8)	NA	3.7
Peripheral neuropathy	154 (15)	73 (10)	2.4	3.4
HFS	109 (10)	65 (9)	2.7	3.6
Neutropenia	42 (4)	209 (28)	4.5	2.4
Lethargy	361 (34)	256 (34)	2.0	2.3
Stomatitis	61 (6)	150 (20)	3.6	2.6
Nausea	210 (20)	88 (12)	2.2	3.1
Vomiting	122 (12)	59 (8)	2.6	3.8
Rash	177 (17)	201 (27)	2.3	2.4

Percentage of patients in parentheses. NA - for neutropenic sepsis in patients treated with XELOX ± cetuximab as there was insufficient power to perform the genome-wide association study. HFS - Hand-foot syndrome

4.2.10 Potential biomarkers of HFS

rs6783836 and potential biomarkers of HFS were analysed using participant data from the UK Biobank. Seven markers of wound healing and/or inflammation were assessed: lymphocyte, neutrophil, monocyte, eosinophil, platelet and basophil counts (10^9 cells/litre) and C-reactive protein levels (mg/l), and one marker for diabetes: glycated haemoglobin (HbA1c) levels (mmol/mol). Analyses were run using PHESANT (Millard *et al*, 2018) (Chapter 2, Section 2.5.4.2). Lymphocyte count, HbA1c levels, platelet count, neutrophil count, c-reactive protein levels and monocyte count were analysed under a linear regression model and, basophil count and eosinophil count were analysed under an ordered logistic model. Results were held to a significance threshold of $P=6.3\text{x}10^{-3}$ (Bonferroni correction for 8 tests, $P=0.05/8$). rs6783836 was analysed as a potential regulator of inflammation by performing a univariate logistic regression on 4,228 individuals from the UK Biobank with self-reported psoriasis and 331,043 controls (as of 1st January 2021).

4.2.11 Additional bioinformatic analyses

The GTEx project database was used to identify eQTLs for relevant SNPs (Chapter 2, Section 2.5.4.2). Significance for tissue association was set at $P<1.0\text{x}10^{-3}$ (i.e. Bonferroni correction for 49 tissues [0.05/49]). Fine mapping was used for SNPs at significant loci using PAINTOR (Kichaev *et al*, 2014). Credible sets of causal SNPs were assembled for 95% coverage. Regional association plots were generated using both the online and JavaScript versions of Locuszoom (Chapter 2, Section 2.4.1).

4.3 Results

4.3.1 Genomic inflation

The distribution of expected and observed *P*-values for each GWAS meta-analysis and their genomic inflation factor (λ range = 0.75 – 0.83) indicated there was deflation of the test statistics. For HFS in patients administered XELOX ± cetuximab, the λ was 0.76. However, individual GWAS test statistics showed no evidence of inflation or deflation, indicating the cause was due to heterogeneity correction in the meta-analysis and not because of abnormal population substructure.

4.3.2 Relationship between genetic variation at *ST6GAL1* and HFS

rs6783836 at 3q27.3 was associated with HFS at genome-wide significant levels in patients treated with XELOX (OR=3.1, 95% CI=2.1-4.6, P=4.3x10^{-8}, **Figure 4.1**). Forty-six percent (50/108) of patients with G2-3 HFS carried rs6783836 in a heterozygous or homozygotes state for the minor allele as compared to 21% (200/934) of patients with G0-1 HFS (**Table 4.3**). The association between rs6783836 and HFS was seen in patients treated with XELOX alone (OR=3.3, 95% CI=1.9-5.7, P=2.7x10^{-5}) and in those treated with XELOX + cetuximab (OR=2.9, 95% CI=1.6-5.1, P=3.0x10^{-4}, **Table 4.3**); cetuximab did not affect this relationship ($P_{interaction}$=0.98). rs6783836 was not associated with HFS in patients treated with FOLFOX (OR=0.86, 95% CI=0.44-1.7, P=0.65) and the difference between regimens was significant ($P_{\text{interaction}}$=1.0x10^{-3}). rs6783836 maps to intron 4 of *ST6GAL1* in a region involved in transcriptional elongation (**Figure 4.2**) and was not an eQTL.

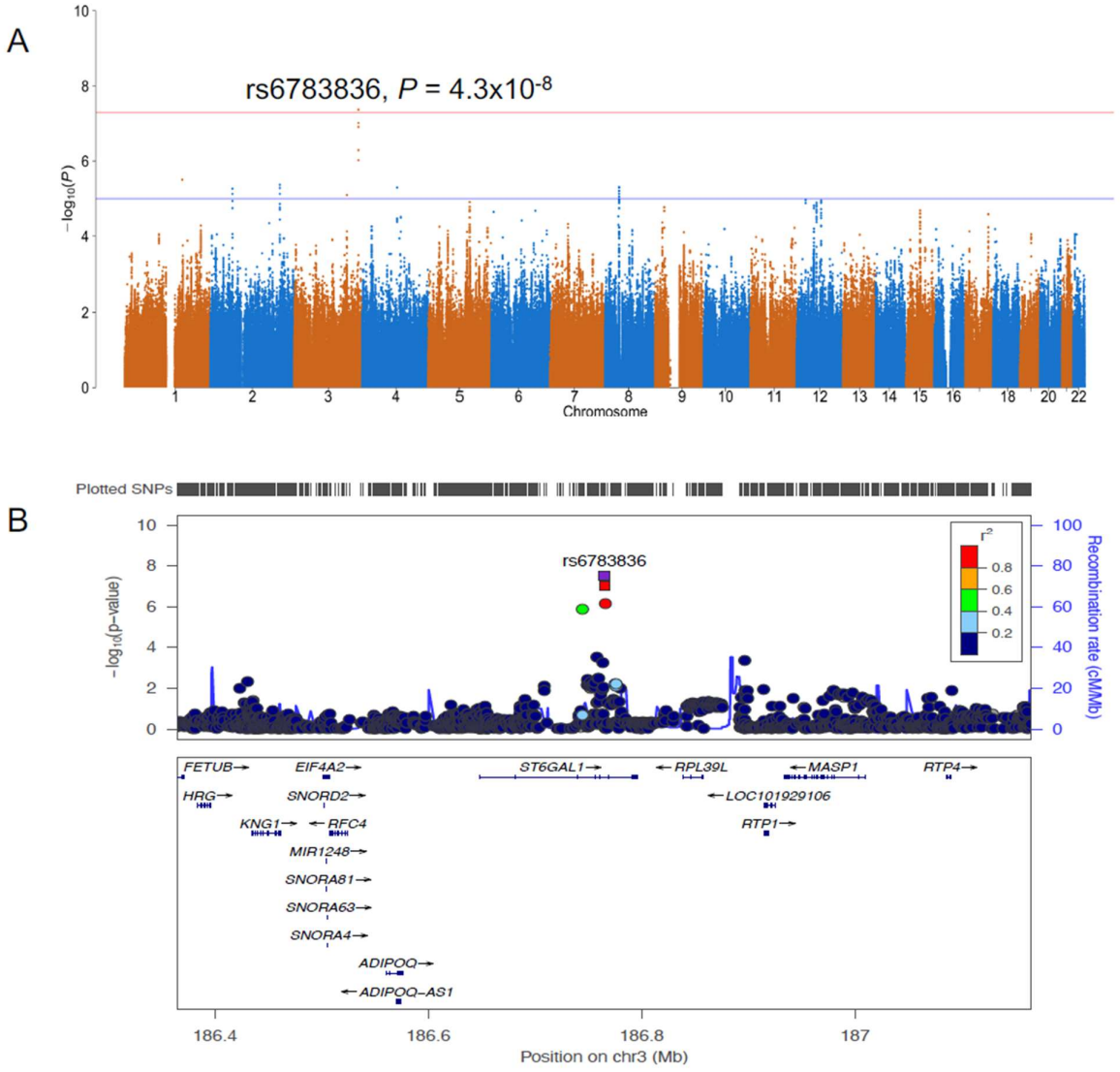

Figure 4.1 Regional plots for the association of rs6783836 with hand-foot syndrome (HFS). (A) Manhattan plot of the association between single-nucleotide polymorphism (SNP) genotype and HFS in patients treated with XELOX. The red line corresponds to a P=5.0x10^{-8} and the blue line P=1.0x10^{-5}. **(B)** Locuszoom plot shows results of the analysis for SNPs and recombination rates. $-\log_{10}(P)$ (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The sentinel SNP (purple) is labelled by its rsID. The colour intensity of each symbol reflects the extent of linkage disequilibrium with the sentinel SNP, deep blue (r^2=0) through to dark red (r^2=1.0). Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a blue line. Physical positions are based on NCBI build 37 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. Fine-mapping identified a credible set of 3 SNPs with rs6783836 having the highest posterior probability of 0.53.

Table 4.3. Relationship between rs6783836 and hand-foot syndrome (HFS) in patients from COIN and COIN-B treated with XELOX ± cetuximab

Treatment groups analysed	Total patients	Patients G0-1 HFS			Patients G2-3 HFS			OR	95% CI	*P*-value
		wildtype	heterozygous	homozygous	wildtype	heterozygous	homozygous			
Meta-analysis	1,042	734	190	10	58	48	2	3.1	2.1-4.6	4.3×10^{-8}
Subgroups:										
XELOX	699	520	121	5	30	21	2	3.3	1.9-5.7	2.7×10^{-5}
XELOX + cetuximab	343	214	69	5	28	27	0	2.9	1.6-5.1	3.0×10^{-4}

Reference allele = T, OR = Odds ratio, CI = Confidence intervals. Eight patients had missing SNP genotyping data for rs6783836.

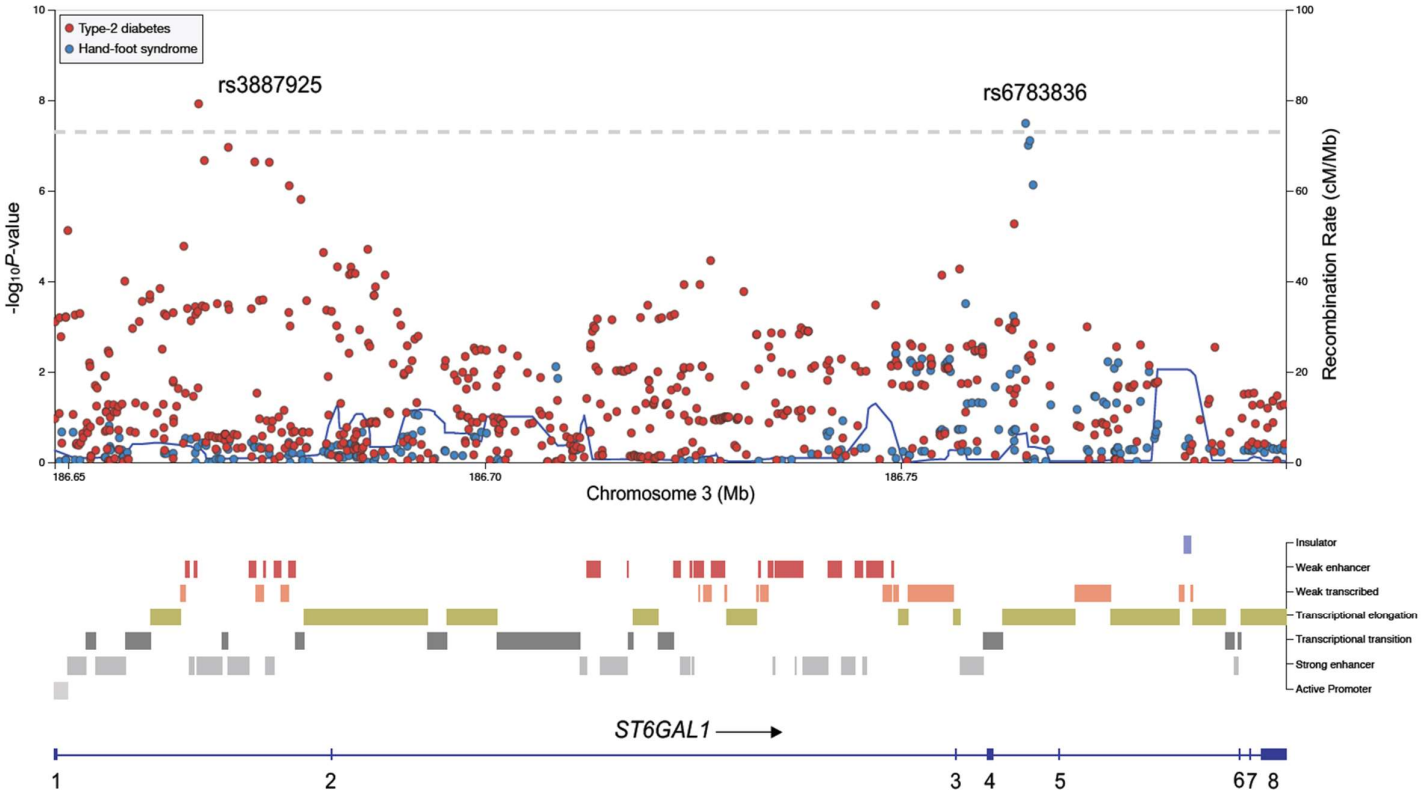

Figure 4.2 Layered Locuszoom plot showing single-nucleotide polymorphisms (SNPs) in *ST6GAL1* associated with hand-foot syndrome (HFS) and type-2 diabetes (T2D). Plot shows results of the analysis for SNPs and recombination rates. $-\log_{10}(P)$ (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The dashed line corresponds to a $P=5.0\times10^{-8}$. Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a blue line. Physical positions are based on NCBI build 37 of the human genome. *ST6GAL1* has previously been associated with T2D, a risk factor for HFS. The association results between SNPs in *ST6GAL1* with HFS (blue), and with T2D (red) are plotted. Lead SNPs for HFS and T2D are indicated by their rsIDs. rs3887925 is not in linkage disequilibrium with rs6783836 (D'=0.26, R^2=0.01), so these appear to be independent loci. Also shown is the relative coding region of *ST6GAL1* and chromatin state annotations from ENCODE.

4.3.3 Investigating the relationship between rs6783836 and HFS in an independent cohort

rs6783836 was borderline significant for HFS in patients treated with only capecitabine from QUASAR2 (OR=0.66, 95% CI=0.42-1.03, *P*=0.05) but with an opposite direction of effect to that found in COIN and COIN-B (**Table 4.4**).

4.3.4 Evaluation of previously purported associations with HFS

Previous studies have shown that common variants in *TYMS/ENSOF1* (rs2612091), *MTHFR* (rs4846048 and rs3737964) and *DPYD* (rs12022243) are associated with HFS. None of these variants replicated in the HFS meta-analysis despite having sufficient power (**Table 4.5**).

Table 4.4 Relationship between rs6783836 and hand-foot syndrome (HFS) in patients from QUASAR2 treated with capecitabine ± bevacizumab

Treatment groups analysed	Total patients	Patients G0-1 HFS			Patients G2-3 HFS			OR	95% CI	*P*-value
		wild type	heterozygous	homozygous	wild type	heterozygous	homozygous			
Capecitabine	440	171	50	1	184	33	1	0.7	0.4-1.0	0.05
Capecitabine + bevacizumab	487	164	35	3	220	57	8	1.3	0.9-1.9	0.24
Meta-analysis	927	335	85	4	404	90	9	0.9	0.7-1.3	0.71

Reference allele = T, OR = Odds ratio, CI = Confidence intervals.

Table 4.5 Lack of replication of loci from previous studies

SNP	Gene	Previous study	COIN and COIN-B	
			XELOX ± cetuximab	FOLFOX ± cetuximab
			***P*-value**	***P*-value**
rs4846048	*MTHFR*	Lin *et al* (2019)	0.64	0.51
rs3737964	*MTHFR*	Lin *et al* (2019)	0.68	0.53
rs2612091	*ENOSF1*	Rosmarin *et al* (2015)	0.73	0.60
rs12022243	*DPYD*	Rosmarin *et al* (2015)	0.40	0.50

4.3.5 Relationship between HFS and patient outcome in COIN and COIN-B

Overall, 174/1,800 (10%) patients from COIN and COIN-B developed G2-3 HFS at 12 weeks (109/1,055, 10% in the XELOX group and 65/745, 9% in the FOLFOX group, **Table 4.2**). HFS was predictive of treatment outcome (**Table 4.6**). 105/154 (68%) patients with G2-3 HFS responded to chemotherapy ± cetuximab at 12 weeks as compared to 831/1436 (58%) with G0-1 HFS (OR =1.6, 95% CI =1.1-2.2, P=1.4x10^{-2}, univariate model). Under a multivariate model accounting for age, sex, disease site, World Health Organisation performance status, primary tumour resection status, white blood cell count, chemotherapy regimen and cetuximab status, this remained significant (OR=1.1, 95% CI=1.02-1.2, P=2.0x10^{-2}). Median OS was 596 days in those with G2-3 HFS and 503 days in those with G0-1 HFS (HR=0.81, 95% CI=0.67-0.97, P=2.4x10^{-2}, **Figure 4.3A**); although, this did not remain significant in multivariate analysis (P=0.15, **Table 4.6**). However, when HFS was assessed as a linear trait, the relationship with OS was significant in both univariate (**Figure 4.3B**) and multivariate analyses (G0 median survival=499 days, G1=514 days, G2=596 days, G3=687 days, HR=0.92, 95% CI=0.84-0.99, P=4.6x10^{-2}, **Table 4.6**).

Table 4.6 Relationship between hand-foot syndrome (HFS) and patient outcome in COIN and COIN-B

Model	Grade of HFS (n)	Response at 12 weeks				Overall survival			
		% Responders	OR	95% CI	*P* (multivariate)	Median survival (days)	HR	95% CI	*P* (multivariate)
Grouped	0-1 (1626)	58	1.6	1.1-2.2	$1.4x10^{-2}$ ($2.0x10^{-2}$)	503	0.81	0.67-0.97	$2.4x10^{-2}$ (0.15)
	2-3 (174)	68				596			
Linear	0 (1264)	56	1.3	1.2-1.6	$1.4x10^{-4}$ ($2.0x10^{-4}$)	499	0.90	0.83-0.97	$5.8x10^{-3}$ ($4.6x10^{-2}$)
	1 (362)	66				514			
	2 (144)	68				596			
	3 (30)	67				687			

Response was defined as complete or partial response using RECIST 1.0 guidelines and no response was defined as stable or progressive disease. 1,800 patients had data on overall survival and 1,590 had data on response at 12 weeks. Covariates included in the multivariate analysis were age, sex, disease site, World Health Organisation performance status, primary tumour resection status, white blood cell count, chemotherapy regimen and cetuximab status. OR = Odds ratio, CI = Confidence intervals, HR = Hazard ratio.

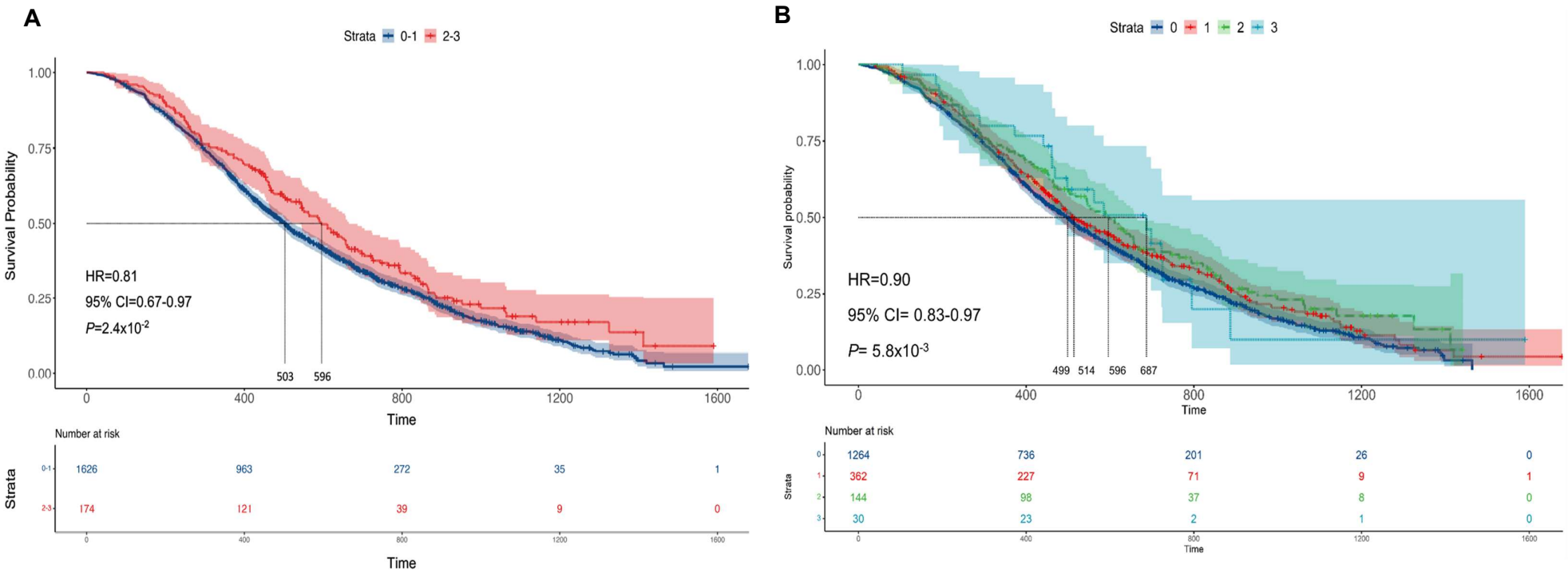

Figure 4.3 Kaplan-Meier plot showing the relationship between hand-foot syndrome (HFS) and overall survival (OS) in patients from COIN and COIN-B, under (A) a grouped model and (B) a linear model. The y-axis represents survival probability and the x-axis represents time (days). Dotted lines show the median OS times **(A)**; 596 days in those with G2-3 HFS and 503 days in those with G0-1 HFS and **(B)**; 499 in those with G0 HFS, 514 in those with G1 HFS, 596 in those with G2 HFS and 687 in those with G3 HFS. The *P*-values were calculated using Cox proportional hazard regression. HR = Hazard ratio, CI = Confidence intervals.

4.3.6 Relationship between rs6783836 in *ST6GAL1* and patient outcome

rs6783836 was not associated with patient outcome regardless of chemotherapy regime (XELOX ± cetuximab, response OR=1.0, 95% CI=0.78-1.4, *P*=0.82 and OS HR=0.95, 95% CI=0.82-1.1, *P*=0.46; FOLFOX ± cetuximab, response OR=0.77, 95% CI=0.54-1.1, *P*=0.15 and OS HR=1.0, 95% CI=0.86-1.2, *P*=0.78).

4.3.7 Understanding the inter-relationship between genetic variation in *ST6GAL1*, T2D and HFS

rs3887925 in intron 1 of *ST6GAL1* was the lead SNP associated with T2D (OR=0.94, 95% CI=0.92-0.96, $P=1.2\times10^{-8}$, **Figure 4.2**), although rs6783836 was not associated with T2D (OR=0.93, 95% CI=0.85-1.0, *P*=0.07) nor diabetic skin lesions (OR=1.1, 95% CI=0.89-1.3, *P*=0.44). rs3887925 and rs6783836 were not in linkage disequilibrium (LD) (D'=0.26, $R^2=0.01$). The rs6783836-T allele was associated with lowered lymphocyte count (beta=-0.0052, 95% CI= -0.0087, -0.0018, $P=2.7\times10^{-3}$) and lowered HbA1c levels (beta= -0.0047, 95% CI= -0.0080, -0.0013, $P=5.9\times10^{-3}$, **Figure 4.4**) that withstood correction for multiple testing. rs6783836 was also associated with psoriasis (OR=0.91, 95% CI=0.85-0.98, $P=7.5\times10^{-3}$).

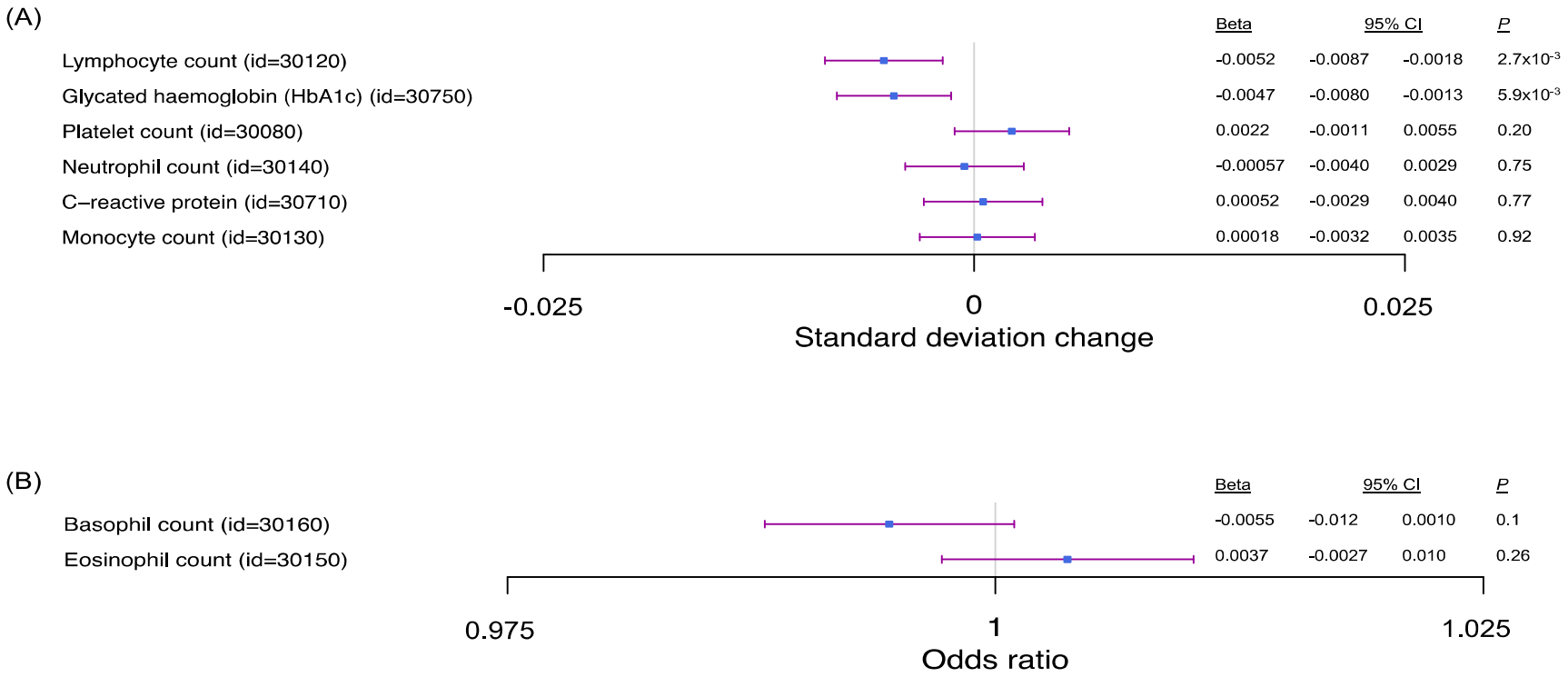

Figure 4.4 Relationship between rs6783836 and (A) continuous and (B) ordinal phenotypes. The x axis shows phenotype and respective UK Biobank ID, and the y axis shows standard deviation change or odds ratio. Only lymphocyte count and glycated haemoglobin (HbA1c) were significantly associated with rs6783836 after Bonferroni correction for 8 tests (P<6.3x10^{-3}). CI = Confidence intervals.

4.3.8 Investigating other variants, genes and pathways associated with toxicities

After meta-analyses, no other SNPs were associated with toxicities in COIN and COIN-B at genome-wide significant levels, but eight SNPs were suggestive of association (P<1.0x10^{-6}). No genes were associated with toxicities after correction for multiple testing (data not shown). Four gene sets (**Table 4.7**) - peripheral neuropathy with response to food, neutropenia with dendritic spine development, diarrhoea with co-receptor activity and skin rash with blood vessel endothelial cell migration, were associated after correction for multiple testing. For the association between peripheral neuropathy with response to food, 5 genes within the set of 20 had P<0.05 (*OXT, SLC16A1, GHRH, G6PC1* and *MPO*).

4.3.9 Investigating toxicity loci identified in Chapter 3

Neither of the genome-wide significant loci identified in Chapter 3, rs13260246 associated with vomiting or *MROH5* associated with neutropenia were significant in XELOX ± cetuximab meta-analyses (rs13260246 OR=3.2, 95% CI=0.8-12.1, P=0.09 and *MROH5* P=0.51). Likewise, none of the 5 suggestive significant loci that replicated in the previous chapter, were significant in these meta-analyses. However, this was expected. For the 2 loci that replicated in the subgroup with the same chemotherapy their association was weakened here due to differences in association strengths (rs9601722 P=1.5x10^{-2}, rs4600090 P=1.1x10^{-3}). The other 3 loci were previously replicated in the subgroup with the same cetuximab status and were not significant in subgroups with the same chemotherapy, hence their non-significance in these meta-analyses (rs6030266 P=0.42, rs1546161 P=0.21, rs13413764 P=0.39).

Table 4.7 MAGMA gene set analyses

Treatment	Toxicity	Pathway	*P*-value
XELOX ± cetuximab	Peripheral neuropathy	Response to food	3.0×10^{-7}
	Neutropenia	Negative regulation of dendritic spine development	5.5×10^{-6}
FOLFOX ± cetuximab	Diarrhoea	Coreceptor activity	2.8×10^{-7}
	Rash	Negative regulation of blood vessel endothelial cell migration	2.7×10^{-6}

Genome-wide significance was a Bonferroni corrected significance threshold of $P<5.6 \times 10^{-6}$.

4.4 Discussion

4.4.1 HFS and treatment efficacy

It has previously been suggested that HFS may be a biomarker of efficacy to chemotherapy (Stintzing *et al*, 2011; Zielinski *et al*, 2016). However, others have suggested that since HFS is a cumulative toxicity, there may be a bias for those living longer simply having more HFS due to having more treatment. This study only considered HFS after 12 weeks of treatment and found that patients with HFS had a better response to chemotherapy at 12 weeks. There was also an improvement in OS when analysed under a linear model of toxicity. Similarly, an exploratory analysis of two German trials noted an association between HFS and OS, and found no difference in OS between patients with early and late HFS (Hofheinz *et al*, 2012). Together, these data suggest that HFS should be tolerated where possible and that an understanding of the underlying mechanism may help improve treatment efficacy.

4.4.2 Exploring the underlying mechanism of rs6783836 in *ST6GAL1*

rs6783836 in *ST6GAL1* was identified as a genome-wide significant biomarker for HFS in patients treated with XELOX, with or without cetuximab. Previous GWAS have revealed that SNPs in *ST6GAL1* are linked to multiple sclerosis (Li *et al*, 2007), coronary artery disease (Saade *et al*, 2011), T2D (Kooner *et al*, 2011), IgA nephropathy (Li *et al*, 2015), asthma (Zhou *et al*, 2019; Oswald *et al*, 2020) and chronic obstructive pulmonary disease (Krick *et al*, 2021). Interestingly, diabetics are at an increased risk of developing HFS, and this study confirmed an association for *ST6GAL1* with T2D. rs6783836 was associated with glycated haemoglobin levels, a marker routinely used in the diagnosis and monitoring of diabetes. However, rs6783836 was not associated with the T2D phenotype directly, with a second locus

at *ST6GAL1* appearing to be causal. It is possible that there is an untagged pleiotropic variant in *ST6GAL1* which is in LD with both the T2D and HFS loci, which would explain the link between these phenotypes. Therefore, further studies in WGS data are required to investigate the LD structure at this locus.

ST6GAL1 also has a known role in inflammation, particularly the NFκB axis signalling pathway through the regulation of TNFR1 (Holdbrooks *et al*, 2020). Circulating ST6GAL1 has been shown to modulate B-cell production (Irons *et al*, 2020), supporting the association between rs6783836 and lymphocyte count. Furthermore, *ST6GAL1* has been associated with psoriasis in an Asian population previously (Wang *et al*, 2017), substantiating the association I identified between psoriasis and rs6783836. It is also interesting to note that *ST6GAL1* knockout mice are viable with mild symptoms including diarrhoea, increased inflammation and defects in B-cell development (Punch *et al*, 2020; Zhang *et al*, 2022).

Overall, although the exact mechanism is unclear, these data support a possible link between HFS, T2D, psoriasis and an underlying defect in the inflammatory pathway, potentially through *ST6GAL1*.

4.4.3 Direction of rs6783836 effect

However, the odds ratios and betas for rs6783836 / *ST6GAL1* with T2D, lymphocyte count, and psoriasis were in the opposite direction to HFS. Lin et al (2007) proposed a flip-flop mechanism for allelic heterogeneity caused by interacting loci in weak LD and this has gained support from recent studies (Wang *et al*, 2018; Maher *et al*, 2010; Zaykin and Shibata, 2008; Shao *et al*, 2016; Mersha *et al*, 2015) and may help

explain our observations. This could occur if an untagged pleiotropic variant is causal for all phenotypes and is in weak LD with both the T2D and HFS loci. Additionally, a main drawback to GWAS is that each variant is considered individually, so more complex genetic effects may appear to ‘flip-flop', despite being genuine (Lin *et al*, 2007). Lin *et al* (2007) demonstrated that the effect direction of one variant can differ between studies due to differences in correlations with other causal variants or environmental factors. Interestingly, the association with HFS was not found in patients treated with FOLFOX and was borderline significant, but with allele flipping, in patients from QUASAR2 treated with capecitabine alone. Other variants in *DPYD* have also demonstrated this flip-flop effect for reasons unknown (Kleibl *et al*, 2009; Joerger *et al*, 2015). While the likelihood of a true flip-flop effect occurring is unlikely, it cannot be discounted altogether at present. Therefore, further studies are needed to investigate this possibility and to understand its potential specificity to particular therapeutic combinations.

4.4.4 Lack of significant variants in other meta-analyses

Despite the increase in statistical power, no significant loci were identified for any other toxicities. There are potential explanations; toxicities could be caused by numerous SNPs with low effect sizes, which were underpowered in these analyses. Alternatively, low-frequency or rare SNPs may explain most phenotypic variance, and these were excluded during QC.

4.4.5 Failure to replicate loci identified in Chapter 3

In the previous chapter (Chapter 3, Section 3.3.4.1), I noted a significant association between rs13260246 and vomiting in patients administered XELOX, but in the meta-

analysis, the association was insignificant. However, this was expected since rs13260246 had failed to replicate in the XELOX + cetuximab subgroup previously. *MROH5* had been significant in both XELOX and XELOX + cetuximab subgroups previously (Chapter 3, Section 3.3.4.8), but the random effects model used for the meta-analyses likely weakened the association due to the heterogeneity between association strengths. The inclusion of covariates during the meta-analysis may also have weakened the association, rendering it insignificant. Therefore, further investigation is needed to explore the relationship between *MROH5* and the covariates associated with neutropenia.

4.4.6 Gene and gene set analyses

Overall, four gene sets were significantly associated with toxicities. Of note, there was an association between peripheral neuropathy and genes involved in response to food which is supported by a previous observation linking diet to chemotherapy-induced peripheral neuropathy (Mongiovi *et al*, 2018). Other forms of peripheral neuropathy have also been linked with diet (Spagnoli *et al*, 2018; Chopra and Tiwari, 2012). Adopting vegetarianism has been shown to relieve symptoms in patients with diabetic neuropathy (Crane and Sample, 1994; Bunner *et al*, 2015), and there is evidence that taking multivitamins reduces the likelihood of a patient experiencing chemotherapy-induced peripheral neuropathy (Zirpoli *et al*, 2017). My data adds weight to this promising avenue for the treatment of this toxicity.

4.4.7 Study limitations

The biggest limitation of this study is the low case rates for some toxicities, even after meta-analysis. This impacts power and so the study was limited to detect only moderate odds ratios.

Additionally, when choosing which subgroups to meta-analyse, I decided to focus on identifying variants associated with XELOX and FOLFOX, at the expense of identifying toxicity variants associated with cetuximab. I decided to do this because XELOX and FOLFOX form the backbone of many different treatment options and are associated with a wide range of toxicities. In comparison, cetuximab is not always administered as a first line treatment and the only notable symptom is often skin rash.

A limitation of the UK Biobank data is that psoriasis was self-reported and thereby prone to misclassification and for wound healing, only indirect measures were available. However, as a large cohort the impact of incorrect classifications was considered minimal.

4.4.8 Conclusions and follow-up studies

This investigation has confirmed an association between HFS and improved patient outcomes. A potential association between *ST6GAL1* and the development of HFS was also identified. However, given the opposite allele effect in the replication cohorts, further replication in larger clinical cohorts is needed to both confirm the initial observation and to explore confounding factors, that may be linked with allele-flipping.

5 Meta-analyses of COIN and QUASAR2 to investigate loci associated with toxicity to capecitabine

5.1 Introduction

5.1.1 Diarrhoea risk factors and mechanisms

Chemotherapy-induced diarrhoea (referred to as diarrhoea *herein*) is one of the most common toxicities, particularly for patients with advanced stage cancer (Stein *et al*, 2010). Upwards of 50% of patients develop diarrhoea at some point during treatment (Maroun *et al*, 2007, Akbarali *et al*, 2022). Diarrhoea is most often associated with fluoropyrimidine chemotherapies, but tyrosine kinase inhibitors and EGFR therapies also increase incidence rates (Benson *et al*, 2004; Leichman *et al*, 1995). Risk factors for diarrhoea include infection, having radiation therapy, prior intestinal resection, being female and having diabetes (Zalcberg *et al*, 1998; Meyerhardt *et al*, 2004). Common genetic variants in *DPYD, ADCY2, MTHFR* and *ABCB1* have also been associated with an increased risk of diarrhoea (Kristensen *et al,* 2010; García-González *et al*, 2015; O'Donnell *et al*, 2020).

The pathophysiology of diarrhoea is complicated and not yet fully understood. There are several documented pathways leading to diarrhoea, working both independently and in combination. At least 5 types of diarrhoea have been described in the literature: (i) secretory, (ii) osmotic, (iii) malabsorption, (iv) exudative and (v) dysmotility.

5.1.1.1 Secretory

The best documented mechanism of diarrhoea is secretory (Thiagarajah *et al*, 2015). During treatment, chemotherapy can damage the intestinal mucosa, leading to a loss of epithelial cells. This reduces the surface area used for water absorption causing higher volumes of fluid to leave the small intestines. This increased fluid output is beyond the absorptive capacity of the colon, resulting in diarrhoea (Keely and Barrett, 2022).

5.1.1.2 Osmotic

Some cancer treatments can disrupt the gastrointestinal osmotic balance which leads to increased fluid output. Similar to secretory diarrhoea, this additional fluid is beyond the absorptive capacity of the colon. EGFR inhibitor induced diarrhoea is a prime example of this mechanism, as it is associated with increased chloride secretion in the gut (Secombe *et al*, 2020; Kim *et al,* 2020; Tao and Chityala, 2021). Osmotic diarrhoea can also be immune-mediated. Chemotherapy can disrupt a patient's microbiome which lets opportunistic infections arise (Ervin *et al*, 2020). Common infections such as *Escherichia coli* then release toxins that can disrupt the gastrointestinal osmotic balance, leading to diarrhoea.

5.1.1.3 Malabsorption

Malabsorption diarrhoea is caused by enzyme deficiencies or changes to substrate absorption, due to chemotherapy damaging intestinal mucosa. The prime example is decreased expression of lactase in the intestinal epithelium which can lead to temporary lactose intolerance (Parnes *et al*, 1994; Österlund *et al*, 2004). Ingestion of milk-containing foods then causes diarrhoea. Chemotherapy-induced bile acid

malabsorption is another common example, and results in shortened colonic transit time, increased water secretion and inhibition of water and electrolyte absorption (Jackson *et al*, 2017).

5.1.1.4 Exudative

Exudative diarrhoea is caused by damage to intestinal mucosa which leads to bleeding and the release of mucosal and submucosal factors into the gut, increasing fluid output (Chassany *et* al, 2012; Field, 2003). Damage to intestinal mucosa also reduces its ability to absorb water amplifying the problem. This damage can occur because of the chemotherapy directly or indirectly due to intestinal inflammation.

5.1.1.5 Dysmotility

Dysmotility is the least understood mechanism of diarrhoea. It is characterised by increased gut motility which reduces the amount of time for fluid absorption causing more fluid to leave the intensities (McQuade *et al*, 2016). A study in mice (Pini *et al*, 2016) showed that cisplatin chemotherapy can cause enteric neuropathy. This resulted in dysmotility diarrhoea due to an increase in the amplitude of neurally induced contractions. Dysmotility as the primary cause of diarrhoea is rare but it is believed to often co-occur with other types (Kroser and Metz, 1996).

5.1.2 Aims

In this chapter, I aimed to identify genetic variants associated with toxicity to capecitabine treatments. I meta-analysed GWAS data for patients in COIN and COIN-B administered XELOX with patients in QUASAR2 (Chapter 1, **Table 1.4**).

This aimed to improve upon my previous analyses, by increasing the power to detect variants with smaller effect sizes.

5.2 Materials and Methods

5.2.1 Patients and samples

As previously described (Chapter 2, Sections 2.2.1.4 and 2.2.1.5), toxicity data and SNP genotypes were available for 1,800 patients from COIN and COIN-B after QC. Out of the 1,800 patients, 707 received XELOX, 345 XELOX + cetuximab, 385 FOLFOX and 360 FOLFOX + cetuximab. XELOX was administered as a 3-week regimen of 130 mg/m^2 intravenous oxaliplatin and 1000 mg/m^2 of oral capecitabine twice daily for 2 weeks, followed by a break of 7 days.

As previously described (Chapter 2, Section 2.2.2), toxicity and SNP genotypes were available for 930 patients from QUASAR2 after QC. Out of the 930 patients, 443 received capecitabine and 487 received capecitabine + bevacizumab. For all patients, capecitabine was administered in 3-week cycles of 1250 mg/m^2 twice daily for 14 days, followed by a break of 7 days.

5.2.2 Toxicities assessed

Toxicities measured in both COIN and QUASAR2 were included in the meta-analysis. These were diarrhoea, HFS, neutropenia, stomatitis and vomiting. In QUASAR2, neutropenia was recorded as haematological toxicity, which also contained a number of patients with platelet toxicity and stomatitis was recorded as mucositis toxicity, which also contained a number of patients with gastrointestinal inflammation. Patients with toxicities graded 2-5 (G2-5) were grouped and compared against those graded 0-1 (G0-1).

5.2.3 Genome-wide association studies

In Chapter 3, four million SNPs were analysed for a relationship with each toxicity under univariate models for each treatment group. Here, I performed similar analyses in Plink (v1.9) but incorporated age and sex as covariates into the additive logistic models to match the covariates used in QUASAR2 (Purcell *et al*, 2007). Analyses were restricted to directly typed SNPs and imputed SNPs with imputation scores ≥0.8, a HWE ≥$1.0x10^{-6}$ and a MAF ≥0.05.

GWAS summary statistics were provided from QUASAR2 analyses by Dr Claire Palles. In their analyses, age and sex were incorporated as covariates. As with COIN, analyses were restricted to directly typed SNPs and imputed SNPs with imputation scores ≥0.8, a HWE ≥$1.0x10^{-6}$ and a MAF ≥0.05.

5.2.4 Meta-analyses

Only the 1,055 patients from COIN administered XELOX ± cetuximab were meta-analysed with QUASAR2, since all these patients were administered capecitabine. All meta-analyses were run under random effects models to account for heterogeneity caused by the differing treatments. Results were plotted in R studio using qqman (Turner, 2018) and ggplot2. SNPs associated at genome-wide significance (P<$5.0x10^{-8}$) were selected for further analyses.

5.2.5 Gene and gene set analyses

MAGMA (de Leeuw *et al*, 2015) was used for gene and gene set analyses. Gene analyses were run under a snpwise univariate model imposing a Bonferroni

corrected significance threshold of P=2.5x10^{-6}. Gene set analyses were run under competitive models with a corrected significance threshold of P=5.8x10^{-6}.

5.2.6 Power considerations

Power to detect toxicity effect sizes was calculated based upon 70% power, P=5.0x10^{-8} and SNPs with MAFs=0.20. Under these conditions, SNPs with a mean OR of 3.5 (range 2.6-5.3 dependent upon toxicity, **Table 5.1**) could be identified.

5.2.7 Validation of rs4760830 in GEL

rs4760830 and potential markers of diarrhoea were analysed using participant data from GEL. For functional diarrhoea, data was extracted from the ICD-10 diagnoses dataset. For the chemotherapy-induced phenotype, cases were classified as patients that had diarrhoea within 3 months of being administered chemotherapy. Diarrhoea data was extracted from the ICD-10 diagnoses dataset, as previously described (Chapter 2 Section 2.2.4.3.1). Chemotherapy data was extracted from the OPCS4 dataset as previously described (Chapter 2 Section 2.2.4.3.2). Only patients administered chemotherapy, but with any cancer, were used in the analyses. Controls were classified as patients administered chemotherapy but that never reported experiencing functional diarrhoea. Overall, 91 participants had functional diarrhoea (28,591 controls) and 16 participants had chemotherapy-induced functional diarrhoea (16,925 controls).

5.2.8 Additional bioinformatic analyses

The GTEx database was used to identify QTLs for relevant SNPs. Significance for tissue association was set at P<1.0x10^{-3} (i.e. Bonferroni correction for 49 tissues

[0.05/49]). Regional association plots were generated using the online version of Locuszoom.

Table 5.1 Meta-analysed patients with grade 2-5 CTCAE toxicities in COIN and QUASAR2 and detectable odds ratios at 70% power

Toxicity	COIN n (%)	QUASAR2 n (%)	Detectable odds ratio
Diarrhoea	288 (27)	199 (21)	2.6
HFS	109 (10)	376 (40)	2.6
Neutropenia	42 (4)	30 (3)	5.3
Stomatitis	61 (6)	69 (7)	4.0
Vomiting	210 (20)	63 (7)	3.0

In COIN, 707 patients were given XELOX (capecitabine + oxaliplatin) and 348 were given XELOX + cetuximab. In QUASAR2, 443 patients were given capecitabine and 487 patients were given capecitabine + bevacizumab. Detectable odds ratios are given for the meta-analysis of COIN and QUASAR2 and calculated for a MAF of 0.05 and *P*-value of $5x10^{-8}$. HFS- hand-foot syndrome.

5.3 Results

5.3.1 Genomic inflation

The distribution of expected and observed *P*-values for each meta-analysed GWAS and their genomic inflation factor (λ range = 0.80-0.81) indicated there was deflation of the test statistics. For diarrhoea in patients administered XELOX ± cetuximab, the λ was 0.80. However, individual GWAS test statistics showed no evidence of inflation or deflation, indicating the cause was due to heterogeneity correction in the meta-analysis and not because of abnormal population substructure.

5.3.2 Relationship between genetic variation at *TRHDE* and diarrhoea

rs4760830 at 12q21.1 was associated with diarrhoea at genome-wide significant levels in patients treated with XELOX or capecitabine (OR=0.6, 95% CI=0.50-0.72, $P=4.8 \times 10^{-8}$, **Figure 5.1**). 76% (1131/1496) of the controls carried rs4760830 in a heterozygous or homozygous state for the minor allele as compared to 66% (322/487) of patients with G2-5 diarrhoea (**Table 5.2**). rs4760830 maps to intron 3 of TRH degrading ectoenzyme (*TRHDE*) and is an eQTL for *TRHDE* in skeletal muscle tissue ($P=2.2 \times 10^{-4}$). In those skeletal muscle samples, the rs4760830 minor allele was associated with increased *TRHDE* gene expression.

5.3.3 Assessment of rs4760830 in patients administered FOLFOX ± cetuximab

rs4760830 was not associated with diarrhoea in patients administered FOLFOX (OR=1.1, 95% CI=0.80-1.6, *P*=0.46) or FOLFOX + cetuximab (OR=1.0, 95% CI=0.70-1.3, *P*=0.83).

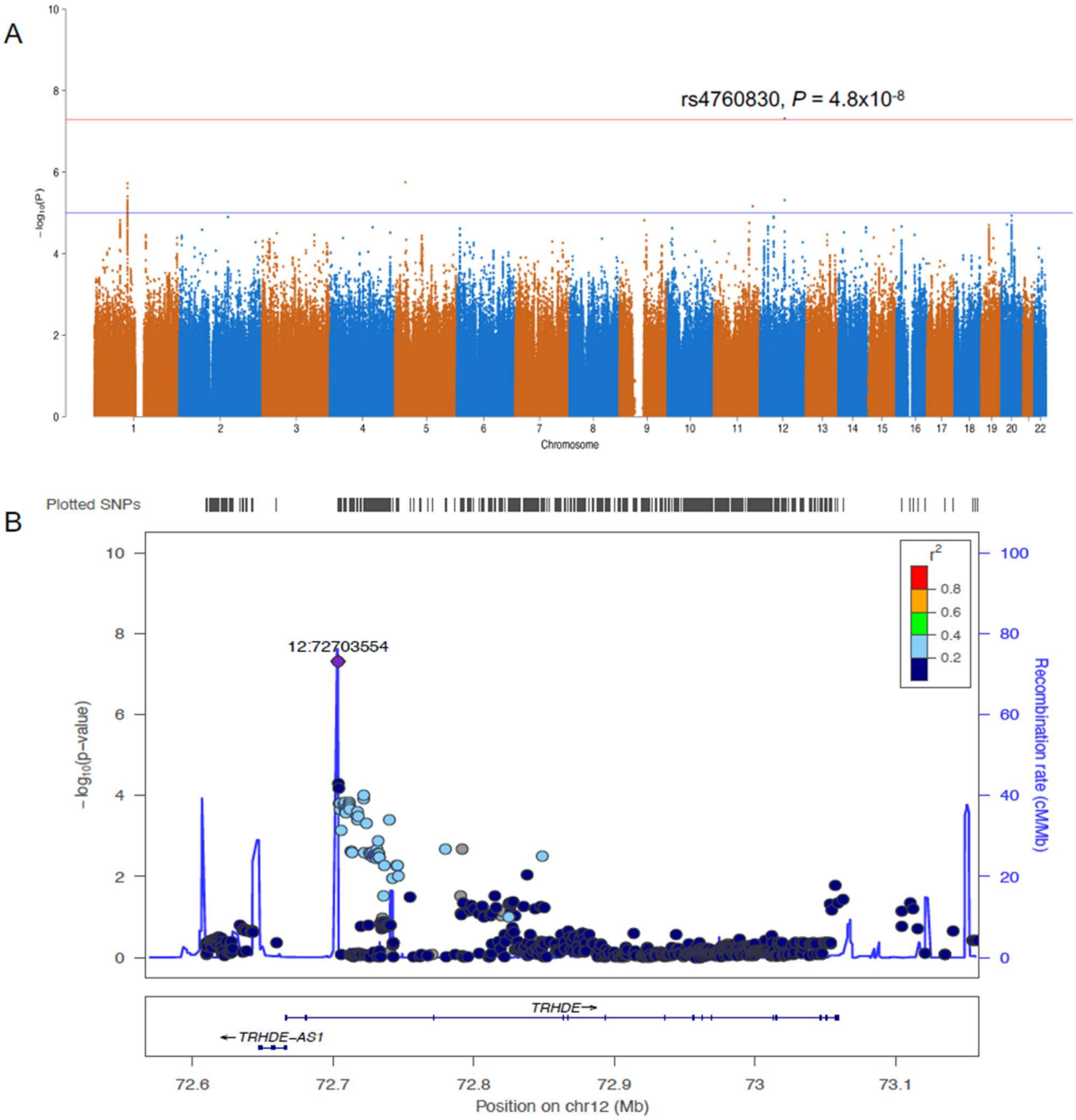

Figure 5.1 Regional plots for the association of rs4760830 with diarrhoea. (A) Manhattan plot of the association between single-nucleotide polymorphism (SNP) genotype and diarrhoea. The red line corresponds to a P=5.0x10^{-8} and the blue line P=1.0x10^{-5}. **(B)** Locuszoom plot shows results of the analysis for SNPs and recombination rates. $-\log_{10}(P)$ (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The sentinel SNP (purple) is labelled by its rsID. The colour intensity of each symbol reflects the extent of linkage disequilibrium with the sentinel SNP, deep blue (r^2=0) through to dark red (r^2=1.0). Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a blue line. Physical positions are based on NCBI build 37 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 5.2 Relationship between rs4760830 genotype and diarrhoea in patients from COIN and QUASAR2 broken down by treatment

Treatment groups analysed	Total patients	Patients G0-1 Diarrhoea			Patients G2-5 Diarrhoea			OR	95% CI	*P*-value
		wild type	heterozygous	homozygous	wild type	heterozygous	homozygous			
Meta-analysis	1983	365	764	367	165	248	74	0.6	0.50-0.72	4.8×10^{-8}
Subgroups:										
XELOX	707	144	287	111	60	80	25	0.7	0.55-0.92	1.0×10^{-2}
XELOX + cetuximab	348	51	113	61	41	63	19	0.6	0.45-0.86	4.1×10^{-3}
Capecitabine	441	79	162	90	35	61	14	0.6	0.44-0.81	1.1×10^{-3}
Capecitabine + bevacizumab	487	91	202	105	29	44	16	0.6	0.44-0.91	1.4×10^{-2}

OR = Odds ratio, CI = Confidence intervals. 2 patients in QUASAR2 in the capecitabine arm had missing data.

5.3.4 Attempted validation of rs4760830 in GEL

rs4760830 failed to associate with functional diarrhoea (OR=0.96, 95% CI=0.7-1.2, *P*=0.77) or chemotherapy-induced diarrhoea (OR=1.1, 95% CI=0.6-2.3, *P*=0.73) in GEL.

5.3.5 Investigating other variants, genes and pathways associated with toxicities

No other SNPs were associated with toxicities in the meta-analyses at genome-wide significant levels, but 21 loci were suggestive of association ($P<1.0x10^{-5}$, **Table 5.3**, not including one SNP that was only present in the two COIN GWAS and therefore excluded from further consideration). Of these, 3 of these were associated with diarrhoea, 7 with HFS, 4 with neutropenia, 6 with stomatitis and 2 with vomiting. No genes or gene sets were associated with toxicities after correction for multiple testing (data not shown).

Table 5.3 Single nucleotide polymorphisms (SNPs) associated with toxicities at P<1.0x10-5 in the meta-analyses

Toxicity	Lead SNP	Cytoband	OR	95% CI	*P*-value
Diarrhoea	rs4760830	12q21.1	0.6	0.50-0.72	$4.8x10^{-8}$
	rs111397431*	5p13.3	6.8	2.3-11.3	$1.9 x10^{-6}$
	rs1356918	1p21.3	1.6	1.3-1.9	$1.9x10^{-6}$
	rs1791807	11q23.2	0.7	0.60-0.82	$6.8x10^{-6}$
HFS	rs1524975	3p14.1	0.6	0.49-0.74	$1.6x10^{-6}$
	rs146460380	6q14.3	2.1	1.5-2.9	$5.6x10^{-6}$
	rs6674251	1q42.2	1.5	1.3-1.8	$6.0x10^{-6}$
	rs2785503	13q33.1	1.5	1.3-1.8	$7.5x10^{-6}$
	rs9615794	22q13.32	1.7	1.4-2.1	$7.6x10^{-6}$
	rs852807	1p32.2	1.4	1.2-1.6	$7.6x10^{-6}$
	rs12603761	17q21.31	1.8	1.4-2.3	$8.1x10^{-6}$
Neutropenia	rs113743917	3p26.3	4.1	2.4-7.0	$3.3x10^{-7}$
	rs73063718	7p15.3	3.0	2.0-4.6	$7.3x10^{-7}$
	rs854034	5q11.2	2.4	1.7-3.4	$1.3x10^{-6}$
	rs12680421	8q24.22	2.4	1.7-3.4	$1.6x10^{-6}$
Stomatitis	rs113336571	6q15	3.1	2.0-4.9	$2.1x10^{-6}$
	rs7670051	4p15.1	2.3	1.6-3.2	$2.7x10^{-6}$
	rs58377730	17q21.31	2.7	1.8-4.0	$2.8x10^{-6}$
	rs2150259	9q31.3	2.2	1.6-3.1	$5.1x10^{-6}$
	8:96825914	8q22.1	0.5	0.37-0.67	$5.5x10^{-6}$
	rs4629011	17q21.32	0.3	0.18-0.50	$6.2x10^{-6}$
Vomiting	rs10002298	4p15.1	2.1	1.5-2.9	$4.8x10^{-6}$
	rs10811964	9p21.3	2.0	1.5-2.7	$9.4x10^{-6}$

HFS- Hand-foot syndrome, OR = Odds ratio, CI = Confidence intervals.
*rs111397431 was present only in COIN and so was excluded from consideration.

5.4 Discussion

5.4.1 Exploring the underlying mechanism of rs4760830 in *TRHDE*

TRHDE is an excellent candidate for diarrhoea. Although *TRHDE* has not previously been linked with diarrhoea it exhibits high expression in neurons of the enteric system (Zeisel *et al*, 2018; May-Zhang *et al*, 2021). The enteric system coordinates gastrointestinal behaviour and dysfunction is often associated with digestive disorders (Rao and Gershon, 2016). As highlighted in Section 5.1.1.5, the enteric system could be affected by chemotherapy and thereby induce dysmotility diarrhoea (Pini *et al*, 2016).

The TRH-DE enzyme has very narrow specificity, displaying strong functional selectivity for TRH (Charli *et al*, 2020). TRH is a peptide primarily expressed in the brain and serum and its best documented function is controlling the hypothalamus–pituitary–thyroid (HPT) axis. However, TRH has also been shown to play a role in the gastrointestinal tract. TRH is linked to secretory and dysmotility diarrhoea, causing marked acceleration of colonic transit by increasing fluid output (Garrick *et* al, 1987; Taché *et al*, 1989). TRH also stimulates gastric secretion and can cause inflammation, both of which can lead to intestinal bleeding and exudative diarrhoea (Taché *et al*, 1989). Moreover, TRH has been implicated in other gastrointestinal disorders. One study found patients with inflammatory bowel disease had a marked accumulation of neuropeptides, including TRH, in the lumen of the colon compared to healthy controls (Yamamoto *et al*, 1996). There are no other enzymes which have documented activity for TRH, so alterations to *TRHDE* would likely have a high impact on TRH activity (Charli *et al*, 2020). The regulation of TRH by *THR-DE* could be an underlying mechanism of diarrhoea. Interestingly, *TRHDE* has an alternative

splicing transcript TRH-DE* which has been shown to regulate TRH-DE enzyme activity (Chavez-Gutierrez *et al*, 2005). Overall, *TRHDE* represents a promising biological candidate but currently lacks genetic validation.

5.4.2 Lack of significant variants in other GWAS

Despite the increase in power from meta-analysing two clinical cohorts, no other significant loci were identified, although 21 loci were suggestive of association. These loci may be potential markers of toxicity, however further validation work is required to support the initial observation, especially due to lack of strong biological evidence for these loci.

The difference in treatments across the meta-analysed groups may be having a confounding effect on results. Whilst the 4 subgroups were given capecitabine, all subgroups had additional treatments that differed from the others. There is also the potential that low-frequency or rare SNPs may play a role, and these were excluded in these analyses.

5.4.3 Study limitations

The biggest limitation of this study is the lack of replication in a capecitabine cohort for the association of rs4760830 and diarrhoea. The association failed to replicate in any of the validation analyses, however the cohorts used could not be filtered for chemotherapy type, and the loci could be capecitabine specific. This is supported by the observation that there was no association between rs4760830 and diarrhoea in patients administered FOLFOX, but the four arms administered capecitabine used in the meta-analysis, all showed nominal significance. Moreover, data on diarrhoea

was limited in GEL, and the only possibly relevant phenotype identified was a severe chronic diarrhoea condition. Therefore, the number of participants used as cases was extremely small, particularly when compared to the high incidence rate for chemotherapy-induced diarrhoea. Therefore, the only conclusion that can be drawn is that rs4760830 is not associated with functional diarrhoea in GEL. rs4760830 may therefore be associated with other diarrhoea phenotypes.

Lastly, a limitation of the meta-analyses was that they were limited to detect only moderate odds ratios so SNPs with low effect sizes (OR <2) may have been missed.

5.4.4 Conclusions and follow up studies

A novel association between a locus in *TRHDE* and the development of diarrhoea was identified. *TRHDE* is a strong candidate gene for diarrhoea and could be clinically useful as diarrhoea is often a dose limiting toxicity for 5FU treatments. The next step would be to identify a cohort which administered capecitabine, to validate the association with rs4760830. This would confirm if the association is treatment specific or a false positive.

6 An analysis of low-frequency SNPs suggests that a nonsynonymous variant in the transporter associated with antigen processing 1 gene predicts chemotherapy-induced sepsis

6.1 Introduction

6.1.1 Low frequency variants

Low frequency variants are SNPs that have MAFs between 1 and 5%. It is well established that rare variants (MAFs less than 1%) are unsuitable for inclusion in GWAS, as specialist genotyping is required to capture these accurately (Weedon *et al*, 2021). However, the inclusion of low-frequency variants is more debatable. For years the standard GWAS cut off has been 5%, with only large population-based studies using a MAF of 1%, due to their increased power capabilities. However, in recent years, more modestly sized GWAS have started to include low-frequency variants to try and explain missing heritability (Manolio *et al*, 2009; Zuk *et al*, 2014). The inclusion of low-frequency variants does however require additional thought in study design (Panoutsopoulou *et al*, 2013). Since the number of individuals with the SNP is low, imputation errors can have a greater effect, thereby giving higher false positive rates compared with analysing common variants (Huang *et al*, 2009). One method of alleviating this problem is to perform independent genotyping of candidate SNPs, to confirm the association is true (Bomba *et al*, 2017).

6.1.2 Neutropenic sepsis

There is no universally accepted definition of neutropenic sepsis (referred to as sepsis from *herein*), but it is characterised by a low neutrophil count and the indication of infection, in patients receiving anticancer treatment (NICE guidelines, 2012b). Sepsis is caused by the immune suppression effect of chemotherapy which allows opportunistic invasive infections to arise (Bhatt and Saleem, 2004). Due to their weakened immune systems, some chemotherapy patients then develop sepsis as a complication. On average, sepsis has a mortality rate upwards of 10% (Klastersky *et al*, 2016), with over 700 cancer patients a year in England and Wales dying from sepsis (Herbst *et al*, 2009).

The diagnosis and treatment of sepsis is complex. The underlying infection could be viral, bacterial, parasitic or fungal in origin which requires different treatment strategies (Freifeld *et al*, 2011; Lin *et al*, 2018). It is also difficult to categorise patients with an infection as being at high or low risk of developing septic complications. Ideally, patients at low risk would be treated with antibiotics as outpatients but patients at high risk would require immediate hospitalisation (Clarke, 2013). Moreover, misclassification has a high impact; admitting all patients would be an unnecessary use of hospital resources but sending a high risk patient home would likely result in their death. Known risk factors for sepsis include age, performance status, chemotherapy type and dosage (NICE guidelines, 2012b). There are currently no known genomic risk factors.

6.1.3 Aims

In previous chapters, I performed GWAS analyses and identified common inherited variants associated with chemotherapy-induced toxicities (Chapters 3-5). Given that low-frequency and rare variants in the gene encoding *DPYD* are associated with toxicities to 5FU (Schwab *et al*, 2008, Henricks *et al*, 2018), in this chapter I investigated the role of low-frequency variants. I then sought related phenotypes and mechanistic understanding using data from the UK Biobank and GEL.

6.2 Materials and Methods

6.2.1 Patients and genotyping

Data on toxicities and SNP genotypes were available for 1,800 patients from COIN and COIN-B after QC (Chapter 2, Sections 2.2.1.4 and 2.2.1.5). Approximately 1.5 million low-frequency SNPs were eligible for analysis after QC. In addition, 16 SNPs were independently genotyped using KASPar technology (LGC, Hertfordshire).

6.2.2 Clinical endpoints

Toxicities assessed were diarrhoea, sepsis, peripheral neuropathy, HFS, neutropenia, lethargy, stomatitis, nausea, vomiting and rash. Patients with toxicities graded 2-5 (G2-5) were grouped and compared against those graded 0-1 (G0-1). For the association between rs56020058 and sepsis, a linear model was also considered to assess if rs56020058 was associated with toxicity severity.

6.2.3 Power considerations

Power to detect SNP effect sizes was calculated, based upon 70% power and P=5.0x10^{-8}. Under these conditions I could identify SNPs with MAFs of 0.05 and 0.01 having mean ORs of 3.0 (2.6-5.4) and 10.1 (8.8-15.9), respectively. For sepsis, the detectable OR were 3.3 and 10.1.

6.2.4 GWAS

GWAS were performed using additive logistic models in Plink (v1.9), incorporating chemotherapy regimen and cetuximab status as covariates. All 1,800 patients with toxicity data were analysed together (**Chapter 1, Table 1.4**). Analyses were restricted to directly typed SNPs and imputed SNPs with imputation scores ≥0.9. A

higher threshold for imputation scores was used here than in other chapters since errors impact low-frequency SNPs more than common SNPs. Other SNP filters were a HWE $\geq 1.0 \times 10^{-6}$ and a MAF ≥0.01 and ≤0.05. SNPs of suggestive significance ($P < 1.0 \times 10^{-5}$) that mapped to or within 50kb of a protein coding gene were selected for additional genotyping. These SNPs were then tested for association again. Only rs56020058 associated with sepsis, remained at a suggestive significance threshold. Results were plotted in R studio using qqman (Turner, 2018).

6.2.5 Bioinformatic analyses

Fine mapping was performed for the *TAP1* locus using PAINTOR (Kichaev *et al*, 2014) (Chapter 2, Section 2.5.4.1). A credible set of causal SNPs was assembled for 95% coverage. Homologous TAP1 sequences were extracted from the NCBI database and aligned using Clustal Omega (Sievers *et al*, 2011). The location of splice sites was predicted using Cytognomix's ValidSpliceMut (Shirley *et al*, 2019). The *in silico* analysis tools CADD (Kircher *et al*, 2014), PolyPhen (Adzhubei *et al*, 2010) and SIFT (Sim *et al*, 2012) were used to predict the impact of variants on protein function.

6.2.6 Sepsis in UK Biobank and GEL

rs1057149 in *TAP1* was analysed for an association with sepsis using participant data from the UK Biobank and GEL. For both, sepsis data was extracted from the ICD-10 diagnoses datasets (Chapter 2 Sections 2.2.3.3.1 and 2.2.4.3.1). Chemotherapy data was extracted from the OPCS4 datasets (Chapter 2 Sections 2.2.3.3.3 and 2.2.4.3.2). Cases were classified as participants that experienced a sepsis event within 3 months of being given chemotherapy. Controls were classified

as participants administered chemotherapy but who had never had sepsis. For CRC phenotypes, data field 40006 in the UK Biobank and the cancer_participant_disease table in GEL were used to filter for participants with a history of CRC or cancer.

In the UK Biobank, 1,091 participants had chemotherapy-induced sepsis (12,090 controls), 92 had CRC chemotherapy-induced sepsis (1,662 controls) and 241 had septic shock (335,030). In GEL, 667 participants had chemotherapy-induced sepsis (3,665), 77 had CRC chemotherapy-induced sepsis (676) and 4,894 had septic shock (30,494).

6.2.7 Immune and inflammatory markers

rs1057149 in *TAP1* was tested for an association with lymphocyte, neutrophil, monocyte, eosinophil, platelet and basophil counts (10^9 cells/litre) in the UK Biobank using PHESANT (Chapter 2 section 2.5.4.2). Lymphocyte, platelet, neutrophil and monocyte counts were analysed under a linear regression model and basophil and eosinophil counts were analysed under an ordered logistic model as software default due to limited variation in the data. Results were held to a significance threshold of P=8.3x10^{-3} (Bonferroni correction for 6 tests, P=0.05/6).

Other loss of function variants in *TAP1* have been associated with chronic respiratory infections. Therefore, I also tested rs1057149 for an association with self-reported septicaemia (326 cases, 334,945 controls), ICD recorded bronchitis (3,016 cases, 332,255 controls), ICD recorded sinusitis (2,232 cases, 333,039 controls) and self-reported sinusitis (1,631 cases, 333,640 controls) in the UK Biobank.

6.3 Results

6.3.1 Genomic inflation

The distribution of expected and observed *P*-values for each GWAS and their genomic inflation factors (λ range= 0.99-1.01) indicated there was no inflation or deflation of the test statistics. Therefore, no underlying population substructure was present.

6.3.2 Association between rs56020058 and sepsis

rs56020058 (imputation score=0.98, MAF in EUR=0.02) in intron 5 of *TAP1* was the most significant SNP associated with sepsis (OR=6.1, 95% CI=3.0-12.8, $P=1.2x10^{-6}$, **Figure 6.1**). Direct genotyping showed 100% (1,733/1,733) genotype concordance with the imputed genotypes. 18.8% (13/69) of patients with G2-5 sepsis carried rs56020058 in a heterozygous state as compared to 4.2% (70/1,686) with G0-1 sepsis (**Table 6.1**). The association remained significant as a linear trait (OR=1.4, 95% CI=1.2-1.6, $P=1.3x10^{-7}$, **Table 6.1**).

Sepsis was more common in patients treated with FOLFOX (8%) than XELOX (1%, $P=2.7x10^{-17}$); however, the relationship between rs56020058 and sepsis was not affected by chemotherapy regimen ($P_{interaction}=0.43$), nor cetuximab status ($P_{interaction}=0.27$) (**Table 6.2**). Neither of the known toxicity-associated *DPYD* variants were associated with sepsis (IVS14+1G>A, 18/1,712 with G0-1 *versus* 2/69 with G2-5 sepsis, *P*=0.08 and Asp949Val, 15/1,714 with G0-1 *versus* 2/69 with G2-5 sepsis, *P*=0.06) and exclusion of the four patients carrying these variants did not significantly affect the association seen with rs56020058 ($P=2.8x10^{-5}$ grouped model).

rs56020058 was also associated with neutropenia (found in 4%, 65/1,511 patients with G0-1 *versus* 8%, 20/247 with G2-5 neutropenia, OR=1.8, 95% CI=1.0-3.2, P=$4.1x10^{-2}$; linear model P=$2.9x10^{-4}$), which itself was associated with sepsis (P=$7.3x10^{-37}$).

rs56020058 was an eQTL for *PSMB8-AS1* (whole blood and skin not sun exposed) and an sQTL for *TAP1* (whole blood and skin not sun exposed) and *PSMB9* (whole blood). The associated *TAP1* isoform was ENST00000486332, which typically has the highest transcripts per million (TPM) in spleen tissue (TPM = 19.5), but also shows some expression in skin not sun exposed tissue (TPM = 8.01). The associated *PSMB9* isoform was ENST00000464863, which typically has low expression in most tissues, being highest also in spleen tissue (TPM = 2.95).

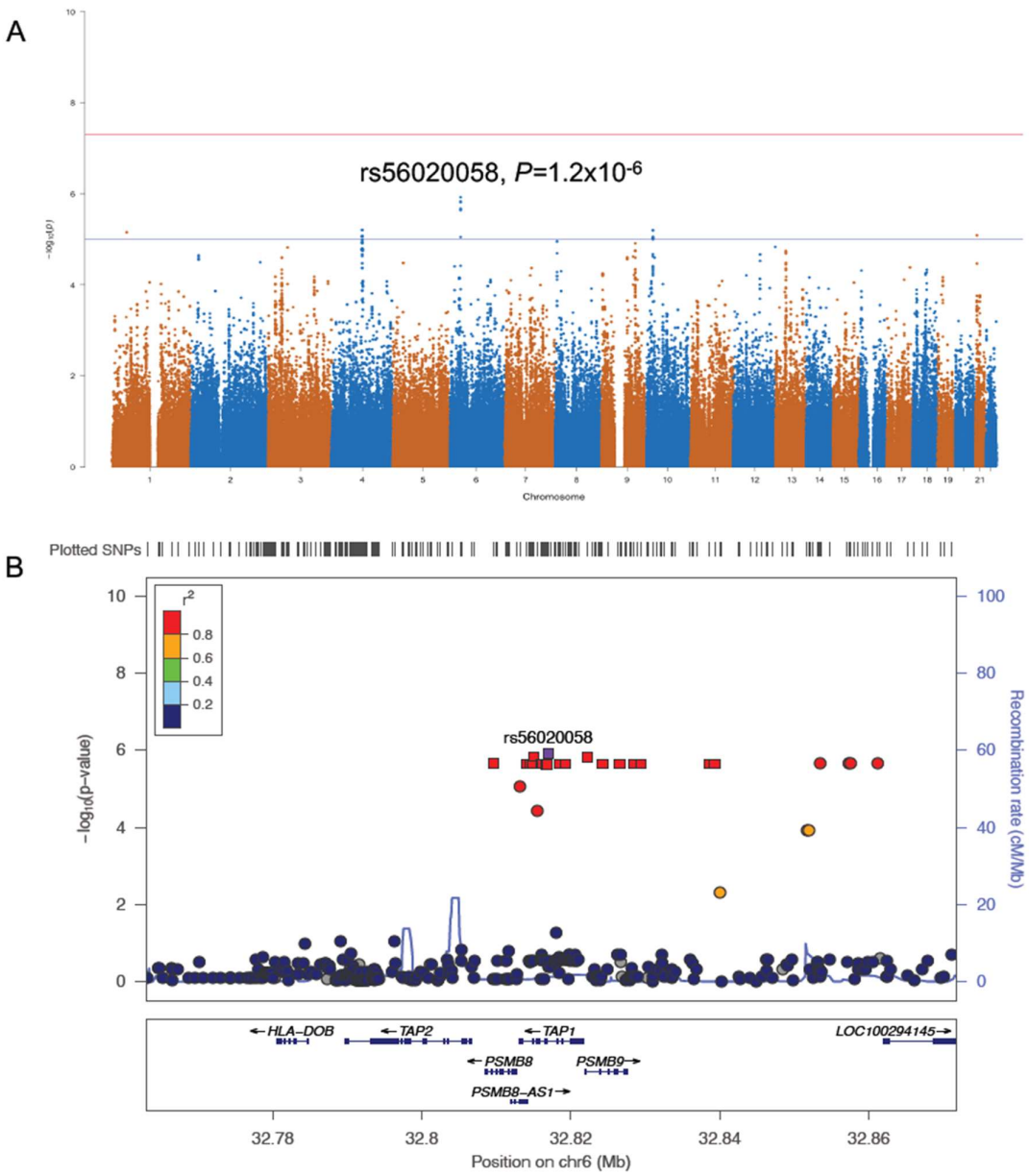

Figure 6.1 Regional plots for the association of rs56020058 with sepsis. (A) Manhattan plot of the association between single-nucleotide polymorphism (SNP) genotype and sepsis. The red line corresponds to a P=5.0x10^{-8} and the blue line P=1.0x10^{-5}. **(B)** Locuszoom plot shows results of the analysis for SNPs and recombination rates. $-\log_{10}(P)$ (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The sentinel SNP (purple) is labelled by its rsID. The colour intensity of each symbol reflects the extent of linkage disequilibrium with the sentinel SNP, deep blue (r^2=0) through to dark red (r^2=1.0). Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a blue line. Physical positions are based on NCBI build 37 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. Fine mapping identified a set of 20 credible SNPs (boxes).

Table 6.1 Relationship between rs56020058 and sepsis

Model	Grade of sepsis	rs56020058 genotype		% of patients	OR	95% CI	*P*-value
		wild type	heterozygous				
Grouped	**0-1**	1616	70	4.2	6.1	3.0-12.8	$1.2x10^{-6}$
	2-5	56	13	18.8			
Linear	**0**	1608	69	4.1	1.4	1.2-1.6	$1.3x10^{-7}$
	1	8	1	11.1			
	2	20	5	20.0			
	3	29	8	21.6			
	4	7	0	0			

OR = Odds ratio, CI = Confidence intervals. No patients were homozygous for the minor allele. 45 patients had missing genotype data. For the linear model the exponential of the beta coefficient is given in place of OR.

Table 6.2 Relationship between rs56020058 and sepsis by treatment

Treatment	Grade of sepsis		OR	95% CI	*P*-value
	G0-1	G2-5			
XELOX	36/702	2/5	13.4	2.2-82.8	$5.4x10^{-3}$
XELOX + cetuximab	14/347	0/1	NA	NA	NA
FOLFOX	13/341	6/24	5.9	1.9-18.5	$2.1x10^{-3}$
FOLFOX + cetuximab	7/296	5/39	4.3	1.3-14.3	$1.8x10^{-2}$

OR – Odds ratio, CI – Confidence intervals, NA – not applicable (insufficient number of cases to run the logistic regression). 45 patients had missing genotype data. The grade of sepsis columns (columns 2 and 3) show the number of patients heterozygous for rs56020058 (no patients were homozygous for the minor allele) out of the total number of patients with that grade of toxicity.

6.3.3 Fine mapping of the *TAP1* locus to identify the causal SNP

Fine mapping identified a 95% credible set of causal variants consisting of 20 SNPs, including two nonsynonymous variants (rs1057149, p.Arg648Gln, c.1943G>A and rs41550019, p.Val458Leu, c.1372G>T) and a potential cryptic acceptor splice site (rs17213826, 1567-159G>T) in *TAP1*, all in linkage disequilibrium with rs56020058 (R^2=1.0, D'=1.0) (**Table 6.3**). rs1057149 was predicted to be dysfunctional by all three *in silico* analysis programmes whereas rs41550019 and rs17213826 were predicted to be benign (**Table 6.3**). rs1057149 lies in *TAP1's* signature region at a moderately conserved residue (**Figure 6.2**).

6.3.4 Attempting to replicate the association in UK Biobank

rs1057149 failed to associate with chemotherapy-induced sepsis in participants with any cancer (OR=0.93, 95% CI=0.68-1.3, *P*=0.62), chemotherapy-induced sepsis in CRC participants (OR=1.29, 95% CI=0.51-3.3, *P*=0.59) or septic shock (OR=1.0, 95% CI=0.54-1.9, *P*=0.99), despite having adequate power (>99%).

6.3.5 Attempting to replicate the association in GEL

rs1057149 failed to associate with chemotherapy-induced sepsis in participants with any cancer (OR=0.21, 95% CI=0.08-0.58, *P*=2.6×10^{-3}), chemotherapy-induced sepsis in CRC patients (OR=0.56, 95% CI=0.07-4.08, *P*=0.57) or septic shock (OR=0.84, 95% CI=0.66-1.07, *P*=0.16), despite having adequate power (>99%).

Table 6.3 Potential causal SNPs in *TAP1* in linkage disequilibrium with rs56020058

SNP	Mutation	Effect	Details	CADD score	PolyPhen score	SIFT	Citations
rs1057149	p.Arg648Gln (c.1943G>A)	Nonsynonymous	Non-conservative amino acid change	35.00	0.98	Deleterious	Yang 2005
rs41550019	p.Val458Leu (c.1372G>T)	Nonsynonymous	Conservative amino acid change	23.70	0.30	Tolerated	NA
rs17213826	1567-159G>T	Intronic	*Potential cryptic acceptor splice site	4.04	NA	NA	NA

NA – not applicable (PolyPhen and SIFT predictions can only be determined for nonsynonymous SNPs). Variants with CADD scores >30 are predicted to be in the top 0.1% of the most deleterious substitutions in the human genome. PolyPhen scores range from 0 (tolerated) to 1 (deleterious) with scores >0.85 predicted to be damaging. *Predicted using ValidSpliceMut

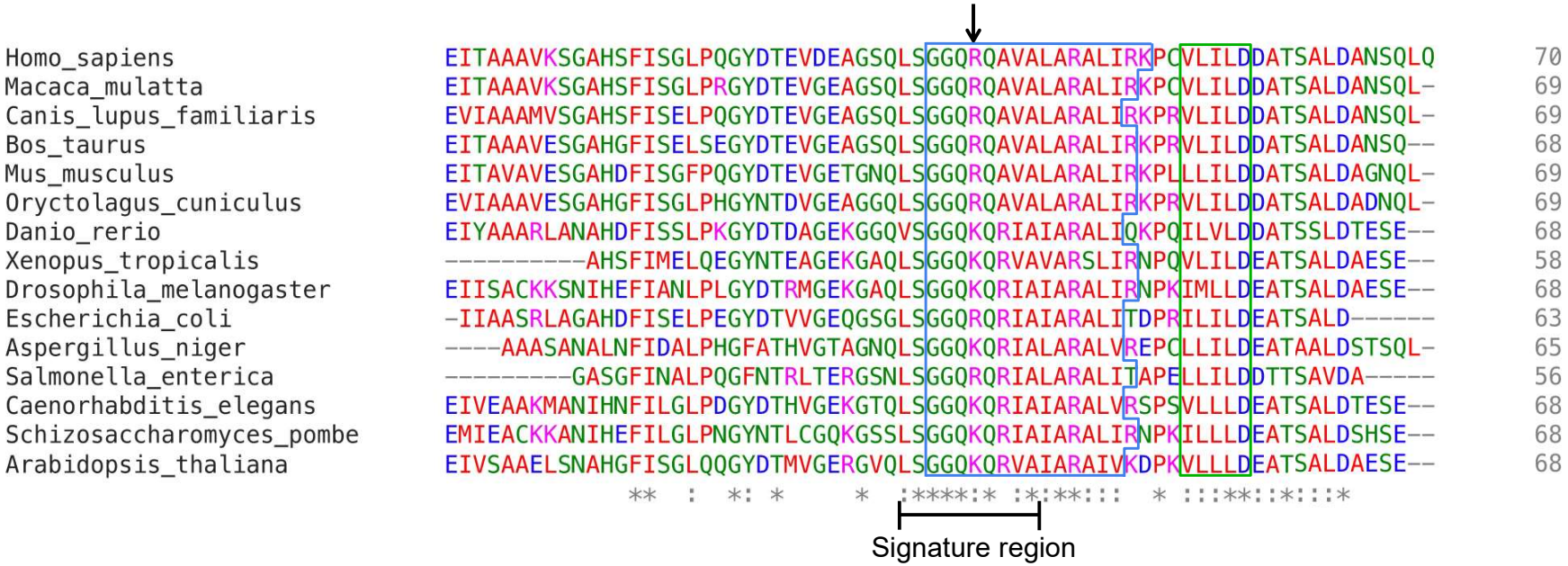

Figure 6.2 Conservation of the TAP1 signature region and rs1057149 (p.Arg648, position 38, arrow, single amino acid code R648). TAP1 homologs - ARB08461.1 (Human), XP_001115506.2 (Rhesus macaque), XP_038538650.1 (Dog), NP_001091527.1 (European cattle), NP_038711.2 (House mouse), XP_002714647.1 (European rabbit), XP_002665053.1 (Zebrafish), XP_002935231.2 (Western clawed frog), NP_523740.3 (Common fruit fly), HBC7235163.1 (Escherichia Coli), GAQ41780.1 (Aspergillus niger), WP_139387570.1 (Salmonella enterica), NP_001300412.1 (Roundworm), BAA01537.1 (Fission yeast) and NP_001327192.1 (Thale cress). Out of 400 organisms with corresponding sequences, 3 (Rattlesnake, Osprey and Wombat) harboured the variant associated with sepsis (R to Q) – normally this residue is conserved with a basic amino acid (R or K). The blue box shows α-helix 5, and the green box shows β-strand 9. * indicates a highly conserved residue, : a moderately conserved residue, and no icon indicates the residue is not conserved.

6.3.6 Association between rs1057149 and immune markers

I sought further evidence for a mechanistic role for rs1057149 in *TAP1* in immune surveillance and found that carriers in the UK Biobank had significantly increased neutrophil (beta=0.060, 95% CI=0.036-0.084, P=1.1x10^{-6}), lymphocyte (beta=0.044, 95% CI=0.023-0.065, P=4.7x10^{-5}) and monocyte (beta=0.0062, 95% CI=0.0023-0.01, P=1.7x10^{-3}) counts and a decreased platelet (beta=-1.8, 95% CI=-0.0032-0.0015, P=6.4x10^{-4}) count (**Table 6.4**). rs1057149 was not associated with septicaemia (P=0.08), bronchitis (P=0.85), or sinusitis (P=0.19).

6.3.7 Other loci of suggestive significance

Four other lead SNPs were suggestive of an association with sepsis. rs1318972 mapping to *BMPR1B*, rs545354772 mapping to *MALRD1*, rs151306999 mapping to *CYP4X1* and rs118000691 mapping to *MIR548XHG* (**Table 6.5**). None of these SNPs were significant QTLs for any genes. For the other toxicities, 11 loci were associated at suggestive significance thresholds. Of these, 1 was associated with peripheral neuropathy (rs33934646), 3 with stomatitis (rs16930421, rs139984076, rs3102166), 1 with nausea (rs1280847118), 4 with vomiting (rs75199447, rs201898372, rs31789, rs77158774), 1 with neutropenia (rs56282256) and 1 with HFS (rs11573121). Direct genotyping showed 93-97% genotype concordance to the imputed genotypes. After re-analysis using the actual genotypes none of these loci remained associated at a suggestive significance threshold, so all were excluded from further analysis.

Table 6.4 Relationship between rs1057149 in *TAP1* and blood immune cells in the UK Biobank

Immune cell type	Beta	95% CI	*P*-value	A-allele effect
Neutrophil	0.059	0.034, 0.03	$1.5x10^{-6}$	Increased
Lymphocyte	0.043	0.022, 0.065	$5.4x10^{-5}$	Increased
Platelets	-1.8	-2.8, - 0.74	$7.1x10^{-4}$	Decreased
Monocytes	0.0060	0.0022, 0.010	$2.0x10^{-3}$	Increased
Basophils	0.0011	0.00023, 0.0020	0.013	NA
Eosinophil	-0.00096	-0.0033, 0.0013	0.42	NA

CI – Confidence Intervals, NA – Not applicable

Table 6.5 Lead SNPs from GWAS associated with sepsis at *P*<1.0x10^{-5}

Lead SNP	Cytoband	Closest gene	OR	95% CI	*P*-value
rs56020058	6p21.32	*TAP1*	6.1	3.0-12.8	1.2x10^{-6}
rs1318972	4q22.3	*BMPR1B*	5.7	2.7-12.2	6.2x10^{-6}
rs545354772	10p12.31	*MALRD1*	8.4	3.3-21.2	6.4x10^{-6}
rs151306999	1p33	*CYP4X1*	9.9	3.6-26.8	7.0x10^{-6}
rs118000691	21q21.1	*MIR548XHG*	9.7	3.6-26.2	8.2x10^{-6}

OR = Odds ratio, CI = Confidence intervals.

6.4 Discussion

6.4.1 Potential clinical utility of rs1057149 in *TAP1*

rs1057149 in *TAP1* has a frequency of 3% in the European population. Carriers of rs1057149 had a greater risk of developing neutropenic sepsis with an effect size (OR=6.1) that has the potential to have clinical impact. In terms of positive predictive value, of the 83 carriers in COIN and COIN-B, 15.7% (n=13) developed G2-5 sepsis. In patients treated with FOLFOX this increased to 35.5% (11/31 carriers developed G2-5 sepsis), although this is less than for the known variants in *DPYD* (IVS14+1G>A, 57.9-73.6% and Asp949Val, 62.5-68.8% for other toxicities in COIN and COIN-B. It is also probable that this initial effect size is inflated due to the winner's curse phenomenon, so the true OR is likely smaller. Therefore if this locus is validated, these validation OR would likely provide a more accurate estimate.

6.4.2 Exploring the underlying mechanism of rs1057149 in *TAP1*

rs1057149 was predicted to affect TAP1 function and lies within the signature region thought to play a role in ATP hydrolysis and peptide transport efficiency (Chen, 2004). A recent study (Chen *et al*, 2022) analysing the effect of human leukocyte antigen (HLA) genes in sepsis using transcriptome data, showed there is lower expression of *TAP1* in all-cause sepsis samples compared to control samples. Another study using colon cancer samples found that the A-allele of rs1057149 occurred 17.5 more frequently in samples with reduced cell surface HLA compared to samples with normal cell surface HLA (Yang *et al*, 2005). Their samples with rs1057149 also showed impaired transporter activity, exhibiting only 20% activity relative to samples with the wild type allele. Together, these data suggest that *TAP1*

and specifically rs1057149 represents a potential biological candidate for sepsis, although further work would be required to confirm the mutational impact.

rs1057149 was also associated with immune blood-based biomarkers levels in the UK Biobank. Interestingly, rs1057149 was associated with a lower platelet count. A study of chronic liver disease patients showed that those with reduced platelet counts, even within normal parameters, had an increased risk of adverse events (Ouyang *et al*, 2021). rs1057149 was also associated with higher neutrophil, lymphocyte and monocyte counts. As discussed in Chapter 4, increased white blood cell counts can be indicative of wound healing and systemic inflammation. Additionally, marginal increases in white blood cell counts can be indicative of various health conditions including T2D (Gkrania-Klotsas *et al*, 2010) and cardiovascular disease (Kabat *et al*, 2017). Patients with rs1057149 may therefore have underlying health conditions that increase their sepsis risk. Loss of function mutations in *TAP1* are causal of bare lymphocyte syndrome, a recessive immunodeficiency disorder where patients experience chronic infections. However, in the UK Biobank, rs1057149 was not associated with chronic respiratory infections.

Despite some supporting biological evidence, the association between rs1057149 and sepsis failed to replicate in any of the validation cohorts. However, no information was available on which chemotherapy was administered, which may have confounded results. Moreover, the initial association failed to reach genome-wide significance, although this was likely due to low power.

6.4.3 Other suggestive significance loci

After re-analysis using the actual genotypes, all other loci were no longer associated at suggestive significance thresholds. It is likely these results were false positives in the initial GWAS, so these were not explored further.

6.4.4 Conclusions and follow-up studies

rs1057149 could be a possible biomarker of sepsis but currently lacks genetic validation, and the initial observation also failed to reach genome-wide significance. Therefore, a priority for future study would be to identify clinical trial cohorts which administered 5FU chemotherapies, that could be used to validate the association. Overall, this study has shown that the analysis of low-frequency variants may help identify novel toxicity biomarkers. However, given how many loci did not remain at even suggestive significance thresholds after genotyping was performed, this highlights the complexities of analysing low-frequency variants in small to medium sized cohorts.

7 General discussion

7.1 Thesis aims

This thesis has focussed on identifying germline biomarkers associated with toxicity to chemotherapeutics. Historically most studies analysing toxicity have used a candidate gene approach to identify potential biomarkers (Rosmarin *et al*, 2014; Pellicer *et al*, 2017). However, this has been hampered by the often-limited understanding of the pathways underlying toxicity (Stein *et al*, 2010; Vichaya *et al*, 2015; Pergolizzi *et al*, 2017). Therefore, this thesis used GWAS methods to identify novel variants and genes associated with toxicity to chemotherapeutics.

7.2 Notable novel findings

7.2.1 Association between *MROH5* and neutropenia

In treatment specific analyses (Chapter 3), the gene *MROH5* was significantly associated with neutropenia in patients treated with XELOX. This association was then independently replicated in those receiving XELOX + cetuximab, suggesting the potential for clinical utility. However, determining whether *MROH5* or the nearby *PTP4A3* is causal is needed before this could be achieved. This is currently limited by the unknown function of *MROH5,* which requires further study.

7.2.2 Association between rs6783836, HFS and inflammation

In the COIN meta-analyses (Chapter 4) I found rs6783836 at *ST6GAL1* was significantly associated with HFS in patients treated with XELOX and was borderline significant in patients receiving capecitabine from QUASAR2, but with an opposite allele effect. In the UK Biobank, *ST6GAL1* was associated with T2D, and the rs6783836-T allele was associated with lowered HbA1c levels, lymphocyte count,

and psoriasis beyond thresholds for multiple testing. Together, these data suggest that inflammatory pathways could be targeted to treat or prevent the development of HFS. This would be of value as I and previous studies have shown that HFS is associated with treatment efficacy (Stintzing *et al*, 2011; Zielinski *et al*, 2016). HFS is often a dose-limiting toxicity for capecitabine, but my data suggest that HFS should be tolerated where possible. However, the issue of allele effect direction must be resolved, before *ST6GAL1* could be useful as a biomarker.

7.2.3 Association between rs4760830 and diarrhoea

In the meta-analyses with QUASAR2 (Chapter 5) I identified that rs4760830 in *TRHDE* was significantly associated with diarrhoea, although the association failed to replicate in GEL. However, rs4760830 was nominally significant in all meta-analysis subgroups, suggesting the potential for replication in other clinical trial cohorts. Furthermore, *TRHDE* is a biologically strong candidate exhibiting high expression in enteric system neurons and its activity impacts several diarrhoea mechanisms (May-Zhang *et al*, 2021; Taché *et al*, 1989; Charli *et al*, 2020). One possibility for the lack of validation in GEL was that the phenotype used was too inaccurate or heterogeneous to show any level of association. Unfortunately, this limitation was due to a lack of well-defined toxicity data available in GEL. Furthermore, neither rs4760830 nor any SNP in strong LD with it were genotyped in UK Biobank. Overall, my results suggest that rs4760830 and *THR-DE* warrant further investigation, especially as diarrhoea is a common side effect of many chemotherapeutics.

7.2.4 Association between *TAP1* locus and sepsis

In the low-frequency GWAS, I identified that rs56020058 in *TAP1* was associated with sepsis (Chapter 6). Although rs56020058 was not genome-wide significant, it represents a strong candidate for chemotherapy-induced sepsis for several reasons. Firstly, it is in LD with a nonsynonymous variant (rs1057149) in *TAP1*. *TAP1* plays a key role in the immune system via the presentation of antigens onto the surface of lymphocytes. rs1057149 is predicted to be dysfunctional by in silico analysis tools and lies in TAP1's signature region at a moderately conserved residue. Previous literature has also shown rs1057149 impairs antigen transporter activity in vitro and is associated with reduced levels of cell surface HLA (Yang *et al*, 2005). Moreover, other rare loss of function mutations in *TAP1* are also causal of Bare Lymphocyte Syndrome type 1, a recessive immunodeficiency disorder characterised by recurrent respiratory infections (Gadola *et al*, 2000; Law-Ping-Man *et al*, 2018). However, the initial observation was not genome-wide significant, and I was biased in following up on this locus due to the strong biological evidence present in the literature for *TAP1.* Furthermore, the locus lacks independent genetic replication at present, with all validation analyses failing to replicate the initial observation.

7.3 Thesis themes and implications of findings

7.3.1 Lack of validation and the potential of markers being therapy specific

One theme throughout this thesis is the lack of validation for investigated loci. On one hand, this could be down to results being false positives, a common problem in GWAS (Kaler and Purcell, 2019). However, there is also the potential that some markers are treatment specific and therefore would not replicate unless treatment

matching cohorts were used. As toxicity incidence rates do differ between XELOX and FOLFOX treatments there must be some underlying difference in the mechanism of toxicity effect, supporting this possibility. Another possibility is that the heterogeneity between discovery and validation cohorts caused by other unaccounted for variables was too significant which could also confound results (Manchia *et al*, 2013; Liu *et al*, 2008).

7.3.2 Lack of common variants associated with toxicities

Throughout this thesis, another prominent theme is the lack of significant common variants with modest to large effect sizes. This is demonstrated by the QQ plots throughout, with lambdas often lower than expected although still within acceptable parameters. To counteract both this and the low power for some toxicities, I investigated loci at suggestive significance to identify any biologically promising but perhaps underpowered loci. However, the majority of these loci also failed to validate. Overall, these results were unexpected which led me to conclude that either common variants have lower effect sizes which require larger cohorts to find them or that low-frequency or rare variants may play a larger role than common variants.

7.3.3 Potential for use of markers in other cancers

5FU based chemotherapeutics are also used for other common cancers such as breast and pancreatic, so it is therefore important to consider if tumour location has any effect on the relationship between SNPs and toxicity. At present, there is little data in the literature on the effect of tumour location on toxicity frequency and toxicity-associated variants. While differences in administration and dosages may have an impact on toxicity frequency between cancers (Hansen et al, 1996), it is

likely variants will still be applicable, with *DPYD* variants used for 5FU guiding regardless of tumour location (Innocenti *et al*, 2020).

7.3.4 Clinical utility

The detection and validation of toxicity biomarkers is an important avenue in the treatment of cancer patients. This knowledge could prevent patient death or limit severe side effects thereby improving patient quality of life. Depending on effect size, the loci identified in this thesis could be utilised in various ways. Loci with large effects such as the *TAP1* locus may be used as biomarkers in the clinic directly if validated, like the currently used *DPYD* variants (Morawska *et al*, 2018). Loci with smaller effect sizes are not clinically useful at present, considering the cost of genotyping and the lack of infrastructure to guide clinicians. However, in the future these variants could be utilised in combination using methods such as polygenic risk scores (Adeyemo *et al*, 2021).

How the variants are utilised in the clinic could also vary depending on the potential severity of a toxicity. Patients at risk could be kept on treatment but with enhanced monitoring, have dose reductions or be administered a different chemotherapy altogether if the risk and potential severity is high (Morawska *et al*, 2018; Hodroj *et al*, 2021). The use of genomic biomarkers also lends itself to one emerging avenue in personalised cancer care which is the shared decision-making policy. The policy allows both patients and clinicians to have input into treatment decisions and has been shown to improve patient quality of life and patient satisfaction (Kashaf and McGill, 2015; Nayak et al, 2017). Implementation of toxicity markers would allow for

effective shared decision making by allowing patients the chance to balance their quality of life and survival outcomes.

Outside the clinic, novel SNPs or genes can also improve the understanding of cellular pathways involved with toxicity and drug reactions (Cirillo *et al*, 2018). This could ultimately help lead to treatments to prevent or treat toxicity. Furthermore, as some toxicities are linked to patient outcomes, these markers may also help to understand and improve treatment efficacy (Stintzing *et al*, 2011).

7.4 Strengths and limitations of this thesis

The methodologies and cohorts used throughout this thesis have their strengths and limitations which may have impacted thesis findings.

7.4.1 Power and sample size

To date, this is the largest published GWAS for toxicity to CRC chemotherapeutics in terms of both sample size and scope. I have performed GWAS for 10 toxicities, some of which have not been assessed in the literature before.

However, one limitation of the cohort was the low event rate for some toxicities which then reduced power. Moreover, in Chapter 6, power was limited due to both the low event rate of sepsis and the fact MAF is also associated with power. Therefore, any genuine variants with low effect sizes will have been underpowered and not reached significance. I attempted to alleviate this by assessing the biological relevance of any loci at suggestive significance.

7.4.2 Phenotype classification

In COIN and COIN-B, toxicity events were recorded by physicians and therefore the phenotypes analysed are precise. However, for some of the phenotypes used in the meta-analyses, QUASAR2 did not have exact matches available. Nevertheless, the phenotypes used instead were clinically relevant and similar, so this was expected to have little impact on the results. An additional problem of using QUASAR2 was that patients were only administered capecitabine rather than XELOX, and half of the patients were also administered bevacizumab. Therefore, this heterogenicity may have impacted results. However, the most clinically useful markers would be those that are predictive regardless of other treatments administered so these results were useful for assessing the generality of potential markers.

I also utilised the population cohorts, UK Biobank and GEL as validation cohorts. While they were useful due to their size and extensive genomic data, their lack of directly coded toxicity data, frequently limited the ability to match phenotypes with accuracy. Instead, I attempted to utilise the ICD diagnoses datasets to extract clinically relevant phenotypes. However, using the diagnoses dataset presented its own limitations. Since 2018, the NHS has used SNOMED CT codes to record patient diagnoses, but both the UK Biobank and GEL convert these to the older ICD codes. During this process, codes may have been incorrectly matched or valuable information may be lost during conversion, since exact matches between the two systems are not always available (Fung and Xu, 2012). Furthermore, for both cohorts, although I could identify patients with cancer that were given chemotherapy, no data on which chemotherapy regimens were administered was available. Therefore, this heterogeneity could have confounded results. However, it was

expected that most CRC patients would have been administered 5FU based therapies since 5FU is the most common treatment for CRC (Gustavsson *et al*, 2015). Nonetheless, other treatments administered in combination could have still confounded results since they would have their own toxicity profile (Braun and Seymour, 2011).

7.5 Future work

7.5.1 Validation of SNPs

The main focus of any future work would be attempting to validate discussed loci in suitable validation cohorts. Ideally, SNPs would be validated in other clinical trial datasets with matching treatment regimens. These trials would also need to have similar clinical measures to use as covariates where needed. At present, no other large CRC trials could be identified that had toxicity and genotyping data available. As mentioned earlier, one possible avenue would be to consider trials for other cancers where 5FU and capecitabine chemotherapies were administered.

Additionally, in upcoming years the UK Biobank plans to release data on chemotherapy regimens. Therefore, a future goal would be to repeat any validation analyses that utilised the UK Biobank while adjusting for treatment regimen. This could potentially reveal effects that were previously masked.

7.5.2 Meta-analysis

Once validation cohorts are identified, another future goal would be to perform further meta-analyses using these, together with COIN and QUASAR2. This would increase power and allow for the identification of SNPs with low effect sizes. This

would help determine if the common disease common variant hypothesis or common disease, rare variant hypothesis, is most applicable for chemotherapy-induced toxicities.

7.5.3 Updating COIN imputation

The reference panel that was used for the original imputation of COIN and COIN-B is now over a decade old. Therefore, a future goal would be to update this imputation. Re-imputing with a newer reference panel or combination of reference panels would allow the inclusion of more SNPs and allow for more accurate capture of low-frequency and rare variants. Combining UK10k and 1000 genomes reference panels has been shown to vastly improve the imputation quality of low-frequency and rare variants in particular. In the Framingham Heart Study (based on 105,796 actual genotypes), imputation with 1000 genomes added 15,245,172 low-frequency and rare variants, whereas imputation with 1000 genomes + UK10k added 21,449,101 of these variants, equating to a 1.4x increase (Chou et al. 2016). Given the associated increases in imputation quality score (from 63% to 76%), it would be of great benefit to re-impute COIN using the larger reference panel and repeat analyses to identify any previously missed loci.

7.5.4 Whole exome and whole genome sequencing

Alternatively, performing whole exome sequencing (WES) or whole genome sequencing (WGS) on the COIN and COIN-B samples would also provide high accuracy and maximise the number of variants available for analysis (Höglund *et al*, 2019). The main limitation for this currently is cost, however in future years both

WES and subsequently WGS are likely to become more affordable and commonplace.

In the short-term WES will be the most attractive option for association studies, as over 85% of known disease-causing mutations lie within exomes, whilst costing significantly less than WGS (Lacey *et al*, 2014). However, efforts to functionally annotate the genome will eventually make WGS more attractive and will likely lead to novel loci and mechanisms being identified (Kim and Wei, 2016). This would be of great benefit for toxicity analyses since many of the underlying mechanisms are currently unknown.

7.5.5 Machine learning approaches

Machine learning (ML) models are algorithms capable of analysing complex data to identify patterns or make predictions. In genomics, ML is an emerging method that could build and improve upon GWAS and post-GWAS methodology (Nicholls *et al*, 2020). Whilst GWAS have proven successful at identifying novel loci, there are limitations to the methodology. One limitation is that SNPs are analysed separately, missing any loci interaction effects (McCarthy *et* al, 2008; Slim *et al*, 2020). Another is the time and resources needed to perform downstream analyses post-GWAS in order to provide functional validation (Nicholls *et al*, 2020). This can be especially demanding when investigating polygenic traits where many loci can be identified in a single GWAS. However, studies have shown ML can successfully be used to address both of these limitations and more. To date, ML models have been used to increase GWAS statistical power (Mieth *et al*, 2016), identify epistatic loci (Leem *et*

al, 2014), perform variant and gene prioritisation post-GWAS (Vitsios and Petrovski, 2020) and to construct polygenic risk models (Pare *et al*, 2017).

An interesting future goal would be to incorporate ML methods to measure the predictive ability of the SNPs identified in this study. ML models can be built in varying complexity depending on the project aim and data available. These range from simple regression models to complex ensemble models or deep learning models. Therefore, I would compare algorithms of different complexities to identify the model with the best predictive ability. The main downside to ML methods is that for complex analyses, large sample sizes are required to achieve adequate power. However, with large genomic datasets such as UK Biobank and GEL available, these could be powerful methods for future variant discovery on the provision that toxicity cases can be classified accurately.

7.6 Outlook

The work in this thesis has identified several promising variants associated with toxicities to chemotherapeutics that warrant further investigation.

References

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, 642–653. https://doi.org/10.1016/S1470-2045(11)70102-4

Adeyemo, A., Balaconis, M.K., Darnes, D.R., Fatumo, S., Granados Moreno, P., Hodonsky, C.J., Inouye, M., Kanai, M., Kato, K., Knoppers, B.M., Lewis, A.C.F., Martin, A.R., McCarthy, M.I., Meyer, M.N., Okada, Y., Richards, J.B., Richter, L., Ripatti, S., Rotimi, C.N., Sanderson, S.C., Sturm, A.C., Verdugo, R.A., Widen, E., Willer, C.J., Wojcik, G.L., Zhou, A., 2021. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nat Med 2021 27:11 27, 1876–1884. https://doi.org/10.1038/s41591-021-01549-6

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat Methods 7, 248–249. https://doi.org/10.1038/NMETH0410-248

Afzal, S., Jensen, S.A., Vainer, B., Vogel, U., Matsen, J.P., Sørensen, J.B., Andersen, P.K., Poulsen, H.E., 2009. MTHFR polymorphisms and 5FU-based adjuvant chemotherapy in colorectal cancer. Annals of oncology 20, 1660–1666. https://doi.org/10.1093/ANNONC/MDP046

Aguado, C., García-Paredes, B., Sotelo, M.J., Sastre, J., Díaz-Rubio, E., 2014. Should capecitabine replace 5-fluorouracil in the first-line treatment of metastatic colorectal cancer? World J Gastroenterol 20, 6092. https://doi.org/10.3748/WJG.V20.I20.6092

Akbarali, H.I., Muchhala, K.H., Jessup, D.K., Cheatham, S., 2022. Chemotherapy induced gastrointestinal toxicities. Adv Cancer Res 155. https://doi.org/10.1016/BS.ACR.2022.02.007

Akdeniz, N., Kaplan, M.A., Uncu, D., İnanç, M., Kaya, S., Dane, F., Küçüköner, M., Demirci, A., Bilici, M., Durnalı, A.G., Koral, L., Şendur, M.A.N., Erol, C., Türkmen, E., Ölmez, Ö.F., Açıkgöz, Ö., Laçin, Ş., Şahinli, H., Urakçı, Z., Işıkdoğan, A., 2021. The comparison of FOLFOX regimens with different doses of 5FU for the adjuvant

treatment of colorectal cancer: a multicenter study. Int J Colorectal Dis 36, 1311–1319. https://doi.org/10.1007/S00384-021-03888-9

Alcindor, T., Beauger, N., 2011. Oxaliplatin: a review in the era of molecularly targeted therapy. Current Oncology 18, 18. https://doi.org/10.3747/CO.V18I1.708

Al-Tassan, N.A., Whiffin, N., Hosking, F.J., Palles, C., Farrington, S.M., Dobbins, S.E., Harris, R., Gorman, M., Tenesa, A., Meyer, B.F., Wakil, S.M., Kinnersley, B., Campbell, H., Martin, L., Smith, C.G., Idziaszczyk, S., Barclay, E., Maughan, T.S., Kaplan, R., Kerr, R., Kerr, D., Buchannan, D.D., Ko Win, A., Hopper, J., Jenkins, M., Lindor, N.M., Newcomb, P.A., Gallinger, S., Conti, D., Schumacher, F., Casey, G., Dunlop, M.G., Tomlinson, I.P., Cheadle, J.P., Houlston, R.S., 2015. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. Sci Rep 5. https://doi.org/10.1038/srep10442

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, 1626–1634. https://doi.org/10.1200/JCO.2007.14.7116

Amstutz, U., Farese, S., Aebi, S., Largíadér, C.R., 2009. Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment. Pharmacogenomics 10, 931–944. https://doi.org/10.2217/PGS.09.28

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, 2343–2351. https://doi.org/10.1056/NEJMOA032709

Andreyev, H.J.N., Davidson, S.E., Gillespie, C., Allum, W.H., Swarbrick, E., 2012. Practice guidance on the management of acute and chronic gastrointestinal problems arising as a result of treatment for cancer. Gut. https://doi.org/10.1136/gutjnl-2011-300563

Andreyev, H.J.N., Lalji, A., Mohammed, K., Muls, A.C.G., Watkins, D., Rao, S., Starling, N., Chau, I., Cruse, S., Pitkaaho, V., Matthews, J., Caley, L., Pittordou, V., Adams, C., Wedlake, L., 2020. The FOCCUS study: a prospective evaluation of the frequency, severity and treatable causes of gastrointestinal symptoms during and after chemotherapy. Supportive Care in Cancer 1–11. https://doi.org/10.1007/s00520-020-05610-x

Argyriou, A.A., Cavaletti, G., Antonacopoulou, A., Genazzani, A.A., Briani, C., Bruna, J., Terrazzino, S., Velasco, R., Alberti, P., Campagnolo, M., Lonardi, S., Cortinovis, D., Cazzaniga, M., Santos, C., Psaromyalou, A., Angelopoulou, A., Kalofonos, H.P., 2013. Voltage-gated sodium channel polymorphisms play a pivotal role in the development of oxaliplatin-induced peripheral neurotoxicity: results from a prospective multicenter study. Cancer 119, 3570–3577. https://doi.org/10.1002/CNCR.28234

Baas, J., Krens, L., Bohringer, S., Mol, L., Punt, C., Guchelaar, H.J., Gelderblom, H., 2018. Genome wide association study to identify predictors for severe skin toxicity in colorectal cancer patients treated with cetuximab. PLoS One 13. https://doi.org/10.1371/JOURNAL.PONE.0208080

Baidoun, F., Elshiwy, K., Elkeraie, Y., Merjaneh, Z., Khoudari, G., Sarmini, M.T., Gad, M., Al-Husseini, M., Saad, A., 2021. Colorectal Cancer Epidemiology: Recent Trends and Impact on Outcomes. Curr Drug Targets 22, 998–1009. https://doi.org/10.2174/1389450121999201117115717

Bailly, C., 2019. Irinotecan: 25 years of cancer treatment. Pharmacol Res 148, 104398. https://doi.org/10.1016/J.PHRS.2019.104398

Battle, A., Mostafavi, S., Zhu, X., Potash, J.B., Weissman, M.M., McCormick, C., Haudenschild, C.D., Beckman, K.B., Shi, J., Mei, R., Urban, A.E., Montgomery, S.B., Levinson, D.F., Koller, D., 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res 24, 14–24. https://doi.org/10.1101/GR.155192.113

Benson, A.B., Ajani, J.A., Catalano, R.B., Engelking, C., Kornblau, S.M., Martenson, J.A., McCallum, R., Mitchell, E.P., O'Dorisio, T.M., Vokes, E.E., Wadler, S., 2004. Recommended guidelines for the treatment of cancer treatment-induced diarrhea. J Clin Oncol 22, 2918–2926. https://doi.org/10.1200/JCO.2004.04.132

Bhatt, V., Saleem, A., 2004. Review: Drug-induced neutropenia--pathophysiology, clinical features, and management. Ann Clin Lab Sci 34, 131–137. http://www.annclinlabsci.org/content/34/2/131.long

Biller, L.H., Schrag, D., 2021. Diagnosis and Treatment of Metastatic Colorectal Cancer: A Review. JAMA 325, 669–685. https://doi.org/10.1001/JAMA.2021.0106

Binefa, G., Rodríguez-Moranta, F., Teule, À., Medina-Hayas, M., 2014. Colorectal cancer: From prevention to personalized medicine. World J Gastroenterol 20, 6786. https://doi.org/10.3748/WJG.V20.I22.6786

Bland, J.M., Altman, D.G., 1995. Calculating correlation coefficients with repeated observations: Part 1--Correlation within subjects. BMJ 310, 446. https://doi.org/10.1136/BMJ.310.6977.446

Bleiberg, H., 1998. Oxaliplatin (L-OHP): a new reality in colorectal cancer. Br J Cancer 77 Suppl 4, 1–3. https://doi.org/10.1038/BJC.1998.427

Blondy, S., David, V., Verdier, M., Mathonnet, M., Perraud, A., Christou, N., 2020. 5-Fluorouracil resistance mechanisms in colorectal cancer: From classical pathways to promising processes. Cancer Sci 111, 3142. https://doi.org/10.1111/CAS.14532

Blumenthal, G.M., Gong, Y., Kehl, K., Mishra-Kalyani, P., Goldberg, K.B., Khozin, S., Kluetz, P.G., Oxnard, G.R., Pazdur, R., 2019. Analysis of time-To-Treatment discontinuation of targeted therapy, immunotherapy, and chemotherapy in clinical trials of patients with non-small-cell lung cancer. Annals of Oncology 30, 830–838. https://doi.org/10.1093/annonc/mdz060

Boisdron-Celle, M., Remaud, G., Traore, S., Poirier, A.L., Gamelin, L., Morel, A., Gamelin, E., 2007. 5-Fluorouracil-related severe toxicity: a comparison of different methods for the pretherapeutic detection of dihydropyrimidine dehydrogenase deficiency. Cancer Lett 249, 271–282. https://doi.org/10.1016/J.CANLET.2006.09.006

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. https://doi.org/10.1093/ANNONC/MDQ632

Bomba, L., Walter, K., Soranzo, N., 2017. The impact of rare and low-frequency genetic variants in common disease. Genome Biology 2017 18:1 18, 1–17. https://doi.org/10.1186/S13059-017-1212-4

Borenstein, M., Hedges, L. v., Higgins, J.P.T., Rothstein, H.R., 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 1, 97–111. https://doi.org/10.1002/JRSM.12

Braun, M.S., Richman, S.D., Thompson, L., Daly, C.L., Meade, A.M., Adlard, J.W., Allan, J.M., Parmar, M.K.B., Quirke, P., Seymour, M.T., 2009. Association of molecular markers with toxicity outcomes in a randomized trial of chemotherapy for advanced colorectal cancer: the FOCUS trial. J Clin Oncol 27, 5519–5528. https://doi.org/10.1200/JCO.2008.21.6283

Braun, M.S., Seymour, M.T., 2011. Balancing the efficacy and toxicity of chemotherapy in colorectal cancer. Ther Adv Med Oncol 3, 43. https://doi.org/10.1177/1758834010388342

Britain, C.M., Holdbrooks, A.T., Anderson, J.C., Willey, C.D., Bellis, S.L., 2018. Sialylation of EGFR by the ST6Gal-I sialyltransferase promotes EGFR activation and resistance to gefitinib-mediated cell death. J Ovarian Res 11. https://doi.org/10.1186/S13048-018-0385-0

Brock, K., Homer, V., Soul, G., Potter, C., Chiuzan, C., Lee, S., 2021. Is more better? An analysis of toxicity and response outcomes from dose-finding clinical trials in cancer. BMC Cancer 2021 21:1 21, 1–18. https://doi.org/10.1186/S12885-021-08440-0

Brody, H., 2015. Colorectal cancer. Nature 2015 521:7551 521, S1–S1. https://doi.org/10.1038/521s1a

Brookes, A.J., 1999. The essence of SNPs. Gene 234, 177–186. https://doi.org/10.1016/S0378-1119(99)00219-X

Bunner, A.E., Wells, C.L., Gonzales, J., Agarwal, U., Bayat, E., Barnard, N.D., 2015. A dietary intervention for chronic diabetic neuropathy pain: A randomized controlled pilot study. Nutr Diabetes 5, e158–e158. https://doi.org/10.1038/nutd.2015.8

Buroker, T.R., O'Connell, M.J., Wieand, H.S., Krook, J.E., Gerstner, J.B., Mailliard, J.A., Schaefer, P.L., Levitt, R., Kardinal, C.G., Gesme, D.H., 1994. Randomized comparison of two schedules of fluorouracil and leucovorin in the treatment of advanced colorectal cancer. J Clin Oncol 12, 14–20. https://doi.org/10.1200/JCO.1994.12.1.14

Bush, W.S., Moore, J.H., 2012. Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 8. https://doi.org/10.1371/JOURNAL.PCBI.1002822

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J., 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018 562:7726 562, 203–209. https://doi.org/10.1038/s41586-018-0579-z

Cassidy, J., Twelves, C., Van Cutsem, E., Hoff, P., Bajetta, E., Boyer, M., Rugat, R., Burger, U., Garin, A., Graeven, U., McKendrick, J., Maroun, J., Marshall, J., Osterwalder, B., Pérez-Manga, G., Rosso, R., Rougier, P., Schilsky, R.L., 2002. First-line oral capecitabine therapy in metastatic colorectal cancer: a favorable safety

profile compared with intravenous5-fluorouracil/leucovorin. Annals of Oncology 13, 566–575. https://doi.org/10.1093/ANNONC/MDF089

Castelloe, J., O'brien, R., 2001. Power and Sample Size Determination for Linear Models. Statistics, Data Analysis, and Data Mining. https://www.lexjansen.com/pharmasug/2001/Proceed/Stats/sp16_castelloe.pdf

Castro-Rojas, C.A., Esparza-Mota, A.R., Hernandez-Cabrera, F., Romero-Diaz, V.J., Gonzalez-Guerrero, J.F., Maldonado-Garza, H., Garcia-Gonzalez, I.S., Buenaventura-Cisneros, S., Sanchez-Lopez, J.Y., Ortiz-Lopez, R., Camacho-Morales, A., Barboza-Quintana, O., Rojas-Martinez, A., 2017. Thymidylate synthase gene variants as predictors of clinical response and toxicity to fluoropyrimidine-based chemotherapy for colorectal cancer. Drug Metab Pers Ther 32, 209–218. https://doi.org/10.1515/dmpt-2017-0028

Charli, J.L., Rodríguez-Rodríguez, A., Hernández-Ortega, K., Cote-Vélez, A., Uribe, R.M., Jaimes-Hoy, L., Joseph-Bravo, P., 2020. The Thyrotropin-Releasing Hormone-Degrading Ectoenzyme, a Therapeutic Target? Front Pharmacol 11, 1. https://doi.org/10.3389/FPHAR.2020.00640

Chassany, O., Michaux, A., Bergmann, J.F., 2012. Drug-Induced Diarrhoea. Drug Safety 2000 22:1 22, 53–72. https://doi.org/10.2165/00002018-200022010-00005

Chávez-Gutiérrez, L., Bourdais, J., Aranda, G., Vargas, M.A., Matta-Camacho, E., Ducancel, F., Segovia, L., Joseph-Bravo, P., Charli, J.L., 2005. A truncated isoform of pyroglutamyl aminopeptidase II produced by exon extension has dominant-negative activity. J Neurochem 92, 807–817. https://doi.org/10.1111/J.1471-4159.2004.02916.X

Chen, M., Abele, R., Tampé, R., 2004. Functional non-equivalence of ATP-binding cassette signature motifs in the transporter associated with antigen processing (TAP). J Biol Chem 279, 46073–46081. https://doi.org/10.1074/JBC.M404042200

Chen, S., Villeneuve, L., Jonker, D., Couture, F., Laverdière, I., Cecchin, E., Innocenti, F., Toffoli, G., Lévesque, E., Guillemette, C., 2015. ABCC5 and ABCG1 polymorphisms predict irinotecan-induced severe toxicity in metastatic colorectal cancer patients. Pharmacogenet Genomics 25, 573–583. https://doi.org/10.1097/FPC.0000000000000168

Chen, Z., Boehnke, M., Wen, X., Mukherjee, B., 2021. Revisiting the genome-wide significance threshold for common variant GWAS. G3 Genes|Genomes|Genetics 11. https://doi.org/10.1093/G3JOURNAL/JKAA056

Chen, Z., Chen, R., Ou, Y., Lu, J., Jiang, Q., Liu, G., Wang, L., Liu, Y., Zhou, Z., Yang, B., Zuo, L., 2022. Construction of an HLA Classifier for Early Diagnosis, Prognosis, and Recognition of Immunosuppression in Sepsis by Multiple Transcriptome Datasets. Front Physiol 13, 875. https://doi.org/10.3389/FPHYS.2022.870657/BIBTEX

Cheng, L., Li, M., Hu, J., Ren, W., Xie, L., Sun, Z.P., Liu, B.R., Xu, G.X., Dong, X.L., Qian, X.P., 2014. UGT1A1*6 polymorphisms are correlated with irinotecan-induced toxicity: a system review and meta-analysis in Asians. Cancer Chemother Pharmacol 73, 551–560. https://doi.org/10.1007/S00280-014-2382-3

Chopra, K., Tiwari, V., 2012. Alcoholic neuropathy: Possible mechanisms and future treatment possibilities. Br J Clin Pharmacol 73, 348–362. https://doi.org/10.1111/j.1365-2125.2011.04111.x

Chou, W.C., Zheng, H.F., Cheng, C.H., Yan, H., Wang, L., Han, F., Richards, J.B., Karasik, D., Kiel, D.P., Hsu, Y.H., 2016. A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples. Scientific Reports 2016 6:1 6, 1–9. https://doi.org/10.1038/srep39313

Cirillo, E., Kutmon, M., Hernandez, M.G., Hooimeijer, T., Adriaens, M.E., Eijssen, L.M.T., Parnell, L.D., Coort, S.L., Evelo, C.T., 2018. From SNPs to pathways: Biological interpretation of type 2 diabetes (T2DM) genome wide association study (GWAS) results. PLoS One 13. https://doi.org/10.1371/JOURNAL.PONE.0193515

Clarke, R.T., Jenyon, T., Parsons, V.V.H., King, A.J., 2013. Neutropenic sepsis: management and complications. Clin Med 13, 185. https://doi.org/10.7861/CLINMEDICINE.13-2-185

Collie-Duguid, E.S.R., Etienne, M.C., Milano, G., McLeod, H.L., 2000. Known variant DPYD alleles do not explain DPD deficiency in cancer patients. Pharmacogenetics 10, 217–223. https://doi.org/10.1097/00008571-200004000-00002

Crane, M.G., Sample, C., 1994. Regression of diabetic neuropathy with total vegetarian (vegan) diet. J Nutr Med 4, 431–439. https://doi.org/10.3109/13590849409003592

Csoboz, B., Gombos, I., Tatrai, E., Tovari, J., Kiss, A.L., Horvath, I., Vigh, L., 2018. Chemotherapy induced PRL3 expression promotes cancer growth via plasma membrane remodeling and specific alterations of caveolae-associated signaling. Cell Communication and Signaling 16, 51. https://doi.org/10.1186/s12964-018-0264-8

Cunningham, D., Humblet, Y., Siena, S., Khayat, D., Bleiberg, H., Santoro, A., Bets, D., Mueser, M., Harstrick, A., Verslype, C., Chau, I., van Cutsem, E., 2004. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med 351, 337–345. https://doi.org/10.1056/NEJMOA033025

Custodio, A., Moreno-Rubio, J., Aparicio, J., Gallego-Plazas, J., Yaya, R., Maurel, J., Higuera, O., Burgos, E., Ramos, D., Calatrava, A., Andrada, E., López, R., Moreno, V., Madero, R., Cejas, P., Feliu, J., 2014. Pharmacogenetic predictors of severe peripheral neuropathy in colon cancer patients treated with oxaliplatin-based adjuvant chemotherapy: a GEMCAD group study. Annals of Oncology 25, 398–403. https://doi.org/10.1093/ANNONC/MDT546

Das, S., Abecasis, G.R., Browning, B.L., 2018. Genotype imputation from large reference panels. Annu Rev Genomics Hum Genet 19, 73–96. https://doi.org/10.1146/ANNUREV-GENOM-083117-021602

de Falco, V., Napolitano, S., Roselló, S., Huerta, M., Cervantes, A., Ciardiello, F., Troiani, T., 2019. How we treat metastatic colorectal cancer. ESMO Open 4, e000813. https://doi.org/10.1136/ESMOOPEN-2020-000813

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, 2938–2947. https://doi.org/10.1200/JCO.2000.18.16.2938

de Leeuw, C.A., Mooij, J.M., Heskes, T., Posthuma, D., 2015. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol 11, e1004219. https://doi.org/10.1371/journal.pcbi.1004219

de Leeuw, C.A., Neale, B.M., Heskes, T., Posthuma, D., 2016. The statistical properties of gene-set analysis. Nat Rev Genet 2016 17:6 17, 353–364. https://doi.org/10.1038/nrg.2016.29

de Mattia, E., Toffoli, G., 2009. C677T and A1298C MTHFR polymorphisms, a challenge for antifolate and fluoropyrimidine-based therapy personalisation. Eur J Cancer 45, 1333–1351. https://doi.org/10.1016/J.EJCA.2008.12.004

Deenen, M.J., Tol, J., Burylo, A.M., Doodeman, V.D., de Boer, A., Vincent, A., Guchelaar, H.J., Smits, P.H.M., Beijnen, J.H., Punt, C.J.A., Schellens, J.H.M., Cats, A., 2011. Relationship between single nucleotide polymorphisms and haplotypes in DPYD and

toxicity and efficacy of capecitabine in advanced colorectal cancer. Clin Cancer Res 17, 3455–3468. https://doi.org/10.1158/1078-0432.CCR-10-2209

Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M., Wallace, M.B., 2019. Colorectal cancer. Lancet 394, 1467–1480. https://doi.org/10.1016/S0140-6736(19)32319-0

DerSimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. Control Clin Trials 7, 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Desmet, F.O., Hamroun, D., Lalande, M., Collod-Bëroud, G., Claustres, M., Béroud, C., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37, e67. https://doi.org/10.1093/NAR/GKP215

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, 5705–5712. https://doi.org/10.1200/JCO.2008.18.0786

di Stefano, A.L., Labussiere, M., Lombardi, G., Eoli, M., Bianchessi, D., Pasqualetti, F., Farina, P., Cuzzubbo, S., Gallego-Perez-Larraya, J., Boisselier, B., Ducray, F., Cheneau, C., Moglia, A., Finocchiaro, G., Marie, Y., Rahimian, A., Hoang-Xuan, K., Delattre, J.Y., Mokhtari, K., Sanson, M., 2015. VEGFA SNP rs2010963 is associated with vascular toxicity in recurrent glioblastomas and longer response to bevacizumab. J Neurooncol 121, 499–504. https://doi.org/10.1007/S11060-014-1677-X

Diasio, R., 2000. Oral DPD-inhibitory fluoropyrimidine drugs - PubMed. Oncology (Williston Park) 14, 19–23. https://pubmed.ncbi.nlm.nih.gov/11098485/

Dobson, R.J., Munroe, P.B., Caulfield, M.J., Saqi, M.A.S., 2006. Predicting deleterious nsSNPs: an analysis of sequence and structural attributes. BMC Bioinformatics 7. https://doi.org/10.1186/1471-2105-7-217

Dorsett, K.A., Marciel, M.P., Hwang, J., Ankenbauer, K.E., Bhalerao, N., Bellis, S.L., 2021. Regulation of ST6GAL1 sialyltransferase expression in cancer cells. Glycobiology 31, 530–539. https://doi.org/10.1093/GLYCOB/CWAA110

Douillard, J.Y., Sobrero, A., Carnaghi, C., Comella, P., Díaz-Rubio, E., Santoro, A., van Cutsem, E., 2003. Metastatic colorectal cancer: integrating irinotecan into combination and sequential chemotherapy. Ann Oncol 14 Suppl 2. https://doi.org/10.1093/ANNONC/MDG723

Dowle, M., Srinivasan A., 2021. data.table: Extension of `data.frame`. https://CRAN.R-project.org/package=data.table

Duarte, H.O., Rodrigues, J.G., Gomes, C., Hensbergen, P.J., Ederveen, A.L.H., de Ru, A.H., Mereiter, S., Polónia, A., Fernandes, E., Ferreira, J.A., van Veelen, P.A., Santos, L.L., Wuhrer, M., Gomes, J., Reis, C.A., 2021. ST6Gal1 targets the ectodomain of ErbB2 in a site-specific manner and regulates gastric cancer cell sensitivity to trastuzumab. Oncogene 40, 3719–3733. https://doi.org/10.1038/s41388-021-01801-w

Ducreux, M., Bennouna, J., Hebbar, M., Ychou, M., Lledo, G., Conroy, T., Adenis, A., Faroux, R., Rebischung, C., Bergougnoux, L., Kockler, L., Douillard, J.Y., 2011. Capecitabine plus oxaliplatin (XELOX) versus 5-fluorouracil/leucovorin plus oxaliplatin (FOLFOX-6) as first-line treatment for metastatic colorectal cancer. Int J Cancer 128, 682–690. https://doi.org/10.1002/IJC.25369

Egger, M., Smith, G.D., Phillips, A.N., 1997. Meta-analysis: principles and procedures. BMJ 315, 1533. https://doi.org/10.1136/BMJ.315.7121.1533

Eichler, H.G., Abadie, E., Breckenridge, A., Flamion, B., Gustafsson, L.L., Leufkens, H., Rowland, M., Schneider, C.K., Bloechl-Daum, B., 2011. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. Nat Rev Drug Discov 10, 495–506. https://doi.org/10.1038/NRD3501

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. Eur J Cancer 45, 228–247. https://doi.org/10.1016/J.EJCA.2008.10.026

Ervin, S.M., Ramanan, S.V., Bhatt, A.P., 2020. Relationship Between the Gut Microbiome and Systemic Chemotherapy. Dig Dis Sci 65, 874–884. https://doi.org/10.1007/S10620-020-06119-3

Escalante, P.I., Quiñones, L.A., Contreras, H.R., 2021. Epithelial-Mesenchymal Transition and MicroRNAs in Colorectal Cancer Chemoresistance to FOLFOX. Pharmaceutics 13, 1–19. https://doi.org/10.3390/PHARMACEUTICS13010075

Falvella, F.S., Cheli, S., Martinetti, A., Mazzali, C., Iacovelli, R., Maggi, C., Gariboldi, M., Pierotti, M.A., di Bartolomeo, M., Sottotetti, E., Mennitto, R., Bossi, I., de Braud, F., Clementi, E., Pietrantonio, F., 2015. DPD and UGT1A1 deficiency in colorectal cancer patients receiving triplet chemotherapy with fluoropyrimidines, oxaliplatin and irinotecan. Br J Clin Pharmacol 80, 581–588. https://doi.org/10.1111/BCP.12631

Farh, K.K.H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., Hatan, M., Carrasco-Alfonso, M.J., Mayer, D., Luckey, C.J., Patsopoulos, N.A., de Jager, P.L., Kuchroo, V.K., Epstein, C.B., Daly, M.J., Hafler, D.A., Bernstein, B.E., 2015. Genetic and Epigenetic Fine-Mapping of Causal Autoimmune Disease Variants. Nature 518, 337. https://doi.org/10.1038/NATURE13835

Fernández-Rozadilla, C., Cazier, J.B., Moreno, V., Crous-Bou, M., Guinó, E., Durán, G., Lamas, M.J., López, R., Candamio, S., Gallardo, E., Paré, L., Baiget, M., Páez, D., López-Fernández, L.A., Cortejoso, L., García, M.I., Bujanda, L., González, D., Gonzalo, V., Rodrigo, L., Reñé, J.M., Jover, R., Brea-Fernández, A., Andreu, M., Bessa, X., Llor, X., Xicola, R., Palles, C., Tomlinson, I., Castellví-Bel, S., Castells, A., Ruiz-Ponte, C., Carracedo, A., 2013. Pharmacogenomics in colorectal cancer: A genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. Pharmacogenomics 13, 209–217. https://doi.org/10.1038/tpj.2012.2

Food and Drug Administration, 2015. Highlights of prescribing information - XELOXDA.https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/020896s037lbl.pdf

Formica, V., Doldo, E., Antonetti, F.R., Nardecchia, A., Ferroni, P., Riondino, S., Morelli, C., Arkenau, H.T., Guadagni, F., Orlandi, A., Roselli, M., 2017. Biological and predictive role of ERCC1 polymorphisms in cancer. Crit Rev Oncol Hematol 111, 133–143. https://doi.org/10.1016/J.CRITREVONC.2017.01.016

Francini, G., Petrioli, R., Lorenzini, L., Mancini, S., Armenio, S., Tanzini, G., Marsili, S., Aquino, A., Marzocca, G., Civitelli, S., Mariani, L., de Sando, D., Bovenga, S., Lorenzi, M., 1994. Folinic acid and 5-fluorouracil as adjuvant chemotherapy in colon cancer. Gastroenterology 106, 899–906. https://doi.org/10.1016/0016-5085(94)90748-X

Fredrikson, M., Hursti, T.J., Steineck, G., Fürst, C.J., Börjesson, S., Peterson, C., 1994. Delayed chemotherapy-induced nausea is augmented by high levels of endogenous noradrenaline. Br J Cancer 70, 642–645. https://doi.org/10.1038/bjc.1994.364

Freifeld, A.G., Bow, E.J., Sepkowitz, K.A., Boeckh, M.J., Ito, J.I., Mullen, C.A., Raad, I.I., Rolston, K. v., Young, J.A.H., Wingard, J.R., 2011. Clinical Practice Guideline for the Use of Antimicrobial Agents in Neutropenic Patients with Cancer: 2010 Update by the Infectious Diseases Society of America. Clinical Infectious Diseases 52, e56–e93. https://doi.org/10.1093/CID/CIR073

Froehlich, T.K., Amstutz, U., Aebi, S., Joerger, M., Largiadèr, C.R., 2015. Clinical importance of risk variants in the dihydropyrimidine dehydrogenase gene for the prediction of early-onset fluoropyrimidine toxicity. Int J Cancer 136, 730–739. https://doi.org/10.1002/IJC.29025

Froelich, M.F., Stintzing, S., Kumbrink, J., Grünewald, T.G.P., Mansmann, U., Heinemann, V., Kirchner, T., Jung, A., Froelich, M.F., Stintzing, S., Kumbrink, J., Grünewald, T.G.P., Mansmann, U., Heinemann, V., Kirchner, T., Jung, A., 2018. The DNA-polymorphism rs849142 is associated with skin toxicity induced by targeted anti-EGFR therapy using cetuximab. Oncotarget 9, 30279–30288. https://doi.org/10.18632/ONCOTARGET.25689

Fujiwara, Y., Chayahara, N., Mukohara, T., Kiyota, N., Tomioka, H., Funakoshi, Y., Minami, H., 2013. Hypothyroidism in patients with colorectal carcinoma treated with fluoropyrimidines. Oncol Rep 30, 1802–1806. https://doi.org/10.3892/or.2013.2644

Fung, K.W., Xu, J., 2012. Synergism between the Mapping Projects from SNOMED CT to ICD-10 and ICD-10-CM. AMIA Annual Symposium Proceedings 2012, 218. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540534/

Gadola, S.D., Moins-Teisserenc, H.T., Trowsdale, J., Gross, W.L., Cerundolo, V., 2000. TAP deficiency syndrome. Clin Exp Immunol 121, 173. https://doi.org/10.1046/J.1365-2249.2000.01264.X

Gamelin, L., Capitain, O., Morel, A., Dumont, A., Traore, S., Anne, L.B., Gilles, S., Boisdron-Celle, M., Gamelin, E., 2007. Predictive Factors of Oxaliplatin Neurotoxicity: The Involvement of the Oxalate Outcome Pathway. Clinical Cancer Research 13, 6359–6368. https://doi.org/10.1158/1078-0432.CCR-07-0660

García-González, X., Cortejoso, L., García, M.I., García-Alfonso, P., Robles, L., Grávalos, C., González-Haba, E., Marta, P., Sanjurjo, M., López-Fernández, L.A., 2015. Variants in CDA and ABCB1 are predictors of capecitabine-related adverse reactions in colorectal cancer. Oncotarget 6, 6422–6430. https://doi.org/10.18632/ONCOTARGET.3289

Garg, M.B., Lincz, L.F., Adler, K., Scorgie, F.E., Ackland, S.P., Sakoff, J.A., 2012. Predicting 5-fluorouracil toxicity in colorectal cancer patients from peripheral blood cell telomere length: a multivariate analysis. Br J Cancer 2012 107:9 107, 1525–1533. https://doi.org/10.1038/bjc.2012.421

Garrick, T., Buack, S., Veiseh, A., Tache, Y., 1987. Thyrotropin-releasing hormone (TRH) acts centrally to stimulate gastric contractility in rats. Life Sci 40, 649–657. https://doi.org/10.1016/0024-3205(87)90266-9

Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F., Guigó, R., 2021. Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. Nat Commun 2021 12:1 12, 1–16. https://doi.org/10.1038/s41467-020-20578-2

Gerrits, A., Li, Y., Tesson, B.M., Bystrykh, L. v., Weersing, E., Ausema, A., Dontje, B., Wang, X., Breitling, R., Jansen, R.C., de Haan, G., 2009. Expression Quantitative Trait Loci Are Highly Sensitive to Cellular Differentiation State. PLoS Genet 5, e1000692. https://doi.org/10.1371/JOURNAL.PGEN.1000692

Giovannucci, E., 2004. Alcohol, one-carbon metabolism, and colorectal cancer: recent insights from molecular studies. J Nutr 134. https://doi.org/10.1093/JN/134.9.2475S

Giovannucci, E., 2002. Modifiable risk factors for colon cancer. Gastroenterol Clin North Am 31, 925–943. https://doi.org/10.1016/S0889-8553(02)00057-2

Gkrania-Klotsas, E., Ye, Z., Cooper, A.J., Sharp, S.J., Luben, R., Biggs, M.L., Chen, L.K., Gokulakrishnan, K., Hanefeld, M., Ingelsson, E., Lai, W.A., Lin, S.Y., Lind, L., Lohsoonthorn, V., Mohan, V., Muscari, A., Nilsson, G., Ohrvik, J., Qiang, J.C., Jenny, N.S., Tamakoshi, K., Temelkova-Kurktschiev, T., Wang, Y.Y., Yajnik, C.S., Zoli, M., Khaw, K.T., Forouhi, N.G., Wareham, N.J., Langenberg, C., 2010. Differential White Blood Cell Count and Type 2 Diabetes: Systematic Review and Meta-Analysis of Cross-Sectional and Prospective Studies. PLoS One 5, e13405. https://doi.org/10.1371/JOURNAL.PONE.0013405

Glass, G. v, 1976. Primary, Secondary, and Meta-Analysis of Research. Educational Researcher 5, 8. https://doi.org/10.2307/1174772

Gonzalez, F.J., Fernandez-Salguaro, P., 1995. Diagnostic analysis, clinical importance and molecular basis of dihydropyrimidine dehydrogenase deficiency. Trends Pharmacol Sci 16, 325–327. https://doi.org/10.1016/S0165-6147(00)89065-3

Gonzalez-Haba, E., García, M.I., Cortejoso, L., López-Lillo, C., Barrueco, N., García-Alfonso, P., Alvarez, S., Jiménez, J.L., Martín, M.L., Muñóz-Fernández, M.A., Sanjurjo, M., Lpez-Fernández, L.A., 2010. ABCB1 gene polymorphisms are associated with adverse reactions in fluoropyrimidine-treated colorectal cancer patients. Pharmacogenomics 11, 1715–1723. https://doi.org/10.2217/PGS.10.159

Göring, H.H.H., Terwilliger, J.D., Blangero, J., 2001. Large Upward Bias in Estimation of Locus-Specific Effects from Genomewide Scans. Am J Hum Genet 69, 1357. https://doi.org/10.1086/324471

Gorlov, I.P., Gorlova, O.Y., Sunyaev, S.R., Spitz, M.R., Amos, C.I., 2008. Shifting Paradigm of Association Studies: Value of Rare Single-Nucleotide Polymorphisms. Am J Hum Genet 82, 100. https://doi.org/10.1016/J.AJHG.2007.09.006

Gross, E., Busse, B., Riemenschneider, M., Neubauer, S., Seck, K., Klein, H.G., Kiechle, M., Lordick, F., Meindl, A., 2008. Strong association of a common dihydropyrimidine dehydrogenase gene polymorphism with fluoropyrimidine-related toxicity in cancer patients. PLoS One 3. https://doi.org/10.1371/JOURNAL.PONE.0004003

Gu, J., Lu, H., Chen, C., Gu, Z., Hu, M., Liu, L., Yu, J., Wei, G., Huo, J., 2021. Diabetes mellitus as a risk factor for chemotherapy-induced peripheral neuropathy: a meta-analysis. Supportive Care in Cancer 29, 7461. https://doi.org/10.1007/S00520-021-06321-7

Guindo-Martínez, M., Amela, R., Bonàs-Guarch, S., Puiggròs, M., Salvoro, C., Miguel-Escalada, I., Carey, C.E., Cole, J.B., Rüeger, S., Atkinson, E., Leong, A., Sanchez, F., Ramon-Cortes, C., Ejarque, J., Palmer, D.S., Kurki, M., Aragam, K., Florez, J.C., Badia, R.M., Mercader, J.M., Torrents, D., 2021. The impact of non-additive genetic associations on age-related complex diseases. Nat Commun 2021 12:1 12, 1–14. https://doi.org/10.1038/s41467-021-21952-4

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. https://doi.org/10.3109/07357907.2015.1104689

Gusella, M., Frigo, A.C., Bolzonella, C., Marinelli, R., Barile, C., Bononi, A., Crepaldi, G., Menon, D., Stievano, L., Toso, S., Pasini, F., Ferrazzi, E., Padrini, R., 2009. Predictors of survival and toxicity in patients on adjuvant therapy with 5-fluorouracil for colorectal cancer. Br J Cancer 100, 1549–1557. https://doi.org/10.1038/SJ.BJC.6605052

Gustavsson, B., Carlsson, G., MacHover, D., Petrelli, N., Roth, A., Schmoll, H.J., Tveit, K.M., Gibson, F., 2015. A review of the evolution of systemic chemotherapy in the management of colorectal cancer. Clin Colorectal Cancer 14, 1–10. https://doi.org/10.1016/j.clcc.2014.11.002

Haller, D.G., Cassidy, J., Clarke, S.J., Cunningham, D., van Cutsem, E., Hoff, P.M., Rothenberg, M.L., Saltz, L.B., Schmoll, H.J., Allegra, C., Bertino, J.R., Douillard,

J.Y., Gustavsson, B.G., Milano, G., O'Connell, M., Rustum, Y., Tabernero, J., Gilberg, F., Sirzén, F., Twelves, C., 2008. Potential regional differences for the tolerability profiles of fluoropyrimidines. J Clin Oncol 26, 2118–2123. https://doi.org/10.1200/JCO.2007.15.2090

Hamzic, S., Kummer, D., Froehlich, T.K., Joerger, M., Aebi, S., Palles, C., Thomlinson, I., Meulendijks, D., Schellens, J.H.M., García-González, X., López-Fernández, L.A., Amstutz, U., Largiadèr, C.R., 2020. Evaluating the role of ENOSF1 and TYMS variants as predictors in fluoropyrimidine-related toxicities: An IPD meta-analysis. Pharmacol Res 152, 104594. https://doi.org/10.1016/J.PHRS.2019.104594

Han, J.Y., Lim, H.S., Eun, S.S., Yoo, Y.K., Yong, H.P., Lee, J.E., Jang, I.J., Dae, H.L., Jin, S.L., 2006. Comprehensive analysis of UGT1A polymorphisms predictive for pharmacokinetics and treatment outcome in patients with non-small-cell lung cancer treated with irinotecan and cisplatin. J Clin Oncol 24, 2237–2244. https://doi.org/10.1200/JCO.2005.03.0239

Han, J.Y., Shin, E.S., Lee, Y.S., Ghang, H.Y., Kim, S.Y., Hwang, J.A., Kim, J.Y., Lee, J.S., 2013. A genome-wide association study for irinotecan-related severe toxicities in patients with advanced non-small-cell lung cancer. Pharmacogenomics J 13, 417–422. https://doi.org/10.1038/TPJ.2012.24

Hansen, R.M., Ryan, L., Anderson, T., Krzywda, B., Quebbeman, E., Benson, A., Haller, D.G., Tormey, D.C., 1996. Phase III study of bolus versus infusion fluorouracil with or without cisplatin in advanced colorectal cancer. J Natl Cancer Inst 88, 668–674. https://doi.org/10.1093/JNCI/88.10.668

Hartmann, K., 2015. Thyroid Disorders in the Oncology Patient. J Adv Pract Oncol 6, 99–106. https://doi.org/10.6004/JADPRO.2015.6.2.2

Hecht, J.R., Mitchell, E., Chidiac, T., Scroggin, C., Hagenstad, C., Spigel, D., Marshall, J., Cohn, A., McCollum, D., Stella, P., Deeter, R., Shahin, S., Amado, R.G., 2009. A randomized phase IIIB trial of chemotherapy, bevacizumab, and panitumumab compared with chemotherapy and bevacizumab alone for metastatic colorectal cancer. J Clin Oncol 27, 672–680. https://doi.org/10.1200/JCO.2008.19.8135

Hecht, J.R., Patnaik, A., Malik, I., Venook, A., Berlin, J., Croghan, G., Wiens, B.L., Visonneau, S., Jerian, S., Meropol, N.J., 2004. ABX-EGF monotherapy in patients (pts) with metastatic colorectal cancer (mCRC): An updated analysis. J Clin Oncol 22, 3511–3511. https://doi.org/10.1200/JCO.2004.22.90140.3511

Henricks, L.M., Lunenburg, C.A.T.C., de Man, F.M., Meulendijks, D., Frederix, G.W.J., Kienhuis, E., Creemers, G.J., Baars, A., Dezentjé, V.O., Imholz, A.L.T., Jeurissen, F.J.F., Portielje, J.E.A., Jansen, R.L.H., Hamberg, P., ten Tije, A.J., Droogendijk, H.J., Koopman, M., Nieboer, P., van de Poel, M.H.W., Mandigers, C.M.P.W., Rosing, H., Beijnen, J.H., Werkhoven, E. van, van Kuilenburg, A.B.P., van Schaik, R.H.N., Mathijssen, R.H.J., Swen, J.J., Gelderblom, H., Cats, A., Guchelaar, H.J., Schellens, J.H.M., 2018. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. Lancet Oncol 19, 1459–1467. https://doi.org/10.1016/S1470-2045(18)30686-7

Henricks, L.M., Opdam, F.L., Beijnen, J.H., Cats, A., Schellens, J.H.M., 2017. DPYD genotype-guided dose individualization to improve patient safety of fluoropyrimidine therapy: Call for a drug label update. Annals of Oncology 28, 2915–2922. https://doi.org/10.1093/annonc/mdx411

Herbst, C., Naumann, F., Kruse, E.B., Monsef, I., Bohlius, J., Schulz, H., Engert, A., 2009. Prophylactic antibiotics or G-CSF for the prevention of infections and improvement of survival in cancer patients undergoing chemotherapy. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.CD007107.PUB2

Hodroj, K., Barthelemy, D., Lega, J.C., Grenet, G., Gagnieu, M.C., Walter, T., Guitton, J., Payen-Gay, L., 2021. Issues and limitations of available biomarkers for fluoropyrimidine-based chemotherapy toxicity, a narrative review of the literature. ESMO Open 6, 100125. https://doi.org/10.1016/J.ESMOOP.2021.100125

Hoff, P.M., Ansari, R., Batist, G., Cox, J., Kocha, W., Kuperminc, M., Maroun, J., Walde, D., Weaver, C., Harrison, E., Burger, H.U., Osterwalder, B., Wong, A.O., Wong, R., 2001. Comparison of oral capecitabine versus intravenous fluorouracil plus leucovorin as first-line treatment in 605 patients with metastatic colorectal cancer: results of a randomized phase III study. J Clin Oncol 19, 2282–2292. https://doi.org/10.1200/JCO.2001.19.8.2282

Hofheinz, R.D., Heinemann, V., von Weikersthal, L.F., Laubender, R.P., Gencer, D., Burkholder, I., Hochhaus, A., Stintzing, S., 2012. Capecitabine-associated hand–foot–skin reaction is an independent clinical predictor of improved survival in patients with colorectal cancer. Br J Cancer 107, 1678. https://doi.org/10.1038/BJC.2012.434

Höglund, J., Rafati, N., Rask-Andersen, M., Enroth, S., Karlsson, T., Ek, W.E., Johansson, Å., 2019. Improved power and precision with whole genome sequencing

data in genome-wide association studies of inflammatory biomarkers. Scientific Reports 2019 9:1 9, 1–14. https://doi.org/10.1038/s41598-019-53111-7

Holdbrooks, A.T., Ankenbauer, K.E., Hwang, J., Bellis, S.L., 2020. Regulation of inflammatory signaling by the ST6Gal-I sialyltransferase. PLoS One 15. https://doi.org/10.1371/JOURNAL.PONE.0241850

Hollander, P. den, Rawls, K., Tsimelzon, A., Shepherd, J., Mazumdar, A., Hill, J., Fuqua, S.A.W., Chang, J.C., Osborne, C.K., Hilsenbeck, S.G., Mills, G.B., Brown, P.H., 2016. Phosphatase PTP4A3 promotes triple-negative breast cancer growth and predicts poor patient survival. Cancer Res 76, 1942–1953. https://doi.org/10.1158/0008-5472.CAN-14-0673

Holmans, P., Green, E.K., Pahwa, J.S., Ferreira, M.A.R., Purcell, S.M., Sklar, P., Owen, M.J., O'Donovan, M.C., Craddock, N., 2009. Gene Ontology Analysis of GWA Study Data Sets Provides Insights into the Biology of Bipolar Disorder. Am J Hum Genet 85, 13–24. https://doi.org/10.1016/j.ajhg.2009.05.011

Hong, E.P., Park, J.W., 2012. Sample Size and Statistical Power Calculation in Genetic Association Studies. Genomics Inform 10, 117. https://doi.org/10.5808/GI.2012.10.2.117

Houreh, M.B., Kalkhajeh, P.G., Niazi, A., Ebrahimi, F., Ebrahimie, E., 2018. SpliceDetector: a software for detection of alternative splicing events in human and model organisms directly from transcript IDs. Sci Rep 8. https://doi.org/10.1038/S41598-018-23245-1

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. https://doi.org/10.1371/JOURNAL.PGEN.1000529

Hsu, H.C., Lapke, N., Chen, S.J., Lu, Y.J., Jhou, R.S., Yeh, C.Y., Tsai, W.S., Hung, H.Y., Hsieh, J.C.H., Yang, T.S., Thiam, T.K., You, J.F., 2018. PTPRT and PTPRD deleterious mutations and deletion predict bevacizumab resistance in metastatic colorectal cancer patients. Cancers (Basel) 10. https://doi.org/10.3390/cancers10090314

Huang, J., Zhang, J., Shi, C., Liu, L., Wei, Y., 2016. Survival, recurrence and toxicity of HNSCC in comparison of a radiotherapy combination with cisplatin versus cetuximab: A meta-analysis. BMC Cancer 16, 1–16. https://doi.org/10.1186/S12885-016-2706-2

Huang, L., Wang, C., Rosenberg, N.A., 2009. The relationship between imputation error and statistical power in genetic association studies in diverse populations. Am J Hum Genet 85, 692–698. https://doi.org/10.1016/J.AJHG.2009.09.017

Huber, C.D., Kim, B.Y., Lohmueller, K.E., 2020. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. PLoS Genet 16. https://doi.org/10.1371/JOURNAL.PGEN.1008827

Huitema, A.D.R., Spaander, M., Mathôt, R.A.A., Tibben, M.M., Holtkamp, M.J., Beijnen, J.H., Rodenhuis, S., 2002. Relationship between exposure and toxicity in high-dose chemotherapy with cyclophosphamide, thiotepa and carboplatin. Annals of Oncology 13, 374–384. https://doi.org/10.1093/annonc/mdf052

Hulshof, E.C., de With, M., de Man, F.M., Creemers, G.J., Deiman, B.A.L.M., Swen, J.J., Houterman, S., Koolen, S.L.W., Bins, S., Thijs, A.M.J., Laven, M.M.J., Hövels, A.M., Luelmo, S.A.C., Houtsma, D., Shulman, K., McLeod, H.L., van Schaik, R.H.N., Guchelaar, H.J., Mathijssen, R.H.J., Gelderblom, H., Deenen, M.J., 2022. UGT1A1 genotype-guided dosing of irinotecan: A prospective safety and cost analysis in poor metaboliser patients. Eur J Cancer 162, 148–157. https://doi.org/10.1016/J.EJCA.2021.12.009

Hunter, J.E., Schmidt, F.L., 2000. Fixed Effects vs. Random Effects Meta-Analysis Models: Implications for Cumulative Research Knowledge. Int J Sel Assess 8, 275–292. https://doi.org/10.1111/1468-2389.00156

Hurria, A., Togawa, K., Mohile, S.G., Owusu, C., Klepin, H.D., Gross, C.P., Lichtman, S.M., Gajra, A., Bhatia, S., Katheria, V., Klapper, S., Hansen, K., Ramani, R., Lachs, M., Wong, F.L., Tew, W.P., 2011. Predicting chemotherapy toxicity in older adults with cancer: A prospective multicenter study, in: J Clin Oncol. American Society of Clinical Oncology, pp. 3457–3465. https://doi.org/10.1200/JCO.2011.34.7625

Hurwitz, H., Fehrenbacher, L., Novotny, W., Cartwright, T., Hainsworth, J., Heim, W., Berlin, J., Baron, A., Griffing, S., Holmgren, E., Ferrara, N., Fyfe, G., Rogers, B., Ross, R., Kabbinavar, F., 2004. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. N Engl J Med 350, 2335–2342. https://doi.org/10.1056/NEJMOA032691

Inada, M., Sato, M., Morita, S., Kitagawa, K., Kawada, K., Mitsuma, A., Sawaki, M., Fujita, K., Ando, Y., 2010. Associations between oxaliplatin-induced peripheral neuropathy and polymorphisms of the ERCC1 and GSTP1 genes. Int J Clin Pharmacol Ther 48, 729–734. https://doi.org/10.5414/CPP48729

Ingrand, I., Defossez, G., Lafay-Chebassier, C., Chavant, F., Ferru, A., Ingrand, P., Pérault-Pochat, M.C., 2020. Serious adverse effects occurring after chemotherapy: A general cancer registry-based incidence survey. Br J Clin Pharmacol 86, 711. https://doi.org/10.1111/BCP.14159

Innocenti, F., Kroetz, D.L., Schuetz, E., Dolan, M.E., Ramírez, J., Relling, M., Chen, P., Das, S., Rosner, G.L., Ratain, M.J., 2009. Comprehensive Pharmacogenetic Analysis of Irinotecan Neutropenia and Pharmacokinetics. J Clin Oncol 27, 2604. https://doi.org/10.1200/JCO.2008.20.6300

Innocenti, F., Mills, S.C., Sanoff, H., Ciccolini, J., Lenz, H.-J., Milano, G., 2020. All You Need to Know About DPYD Genetic Testing for Patients Treated With Fluorouracil and Capecitabine: A Practitioner-Friendly Guide . JCO Oncol Pract 16, 793–798. https://doi.org/10.1200/op.20.00553

Innocenti, F., Undevia, S.D., Iyer, L., Chen, P.X., Das, S., Kocherginsky, M., Karrison, T., Janisch, L., Ramírez, J., Rudin, C.M., Vokes, E.E., Ratain, M.J., 2004. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. J Clin Oncol 22, 1382–1388. https://doi.org/10.1200/JCO.2004.07.173

Ioannidis, J.P.A., 2007. Non-Replication and Inconsistency in the Genome-Wide Association Setting. Hum Hered 64, 203–213. https://doi.org/10.1159/000103512

Irons, E.E., Punch, P.R., Lau, J.T.Y., 2020. Blood-Borne ST6GAL1 Regulates Immunoglobulin Production in B Cells. Front Immunol 11, 617. https://doi.org/10.3389/FIMMU.2020.00617/BIBTEX

Iyer, L., Das, S., Janisch, L., Wen, M., Ramírez, J., Karrison, T., Fleming, G.F., Vokes, E.E., Schilsky, R.L., Ratain, M.J., 2002. UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. Pharmacogenomics J 2, 43–47. https://doi.org/10.1038/SJ.TPJ.6500072

Jackson, A., Lalji, A., Kabir, M., Muls, A., Gee, C., Vyoral, S., Shaw, C., Andreyev, H.J.N., 2017. The efficacy of a low-fat diet to manage the symptoms of bile acid malabsorption – outcomes in patients previously treated for cancer. Clin Med 17, 412–418. https://doi.org/10.7861/CLINMEDICINE.17-5-412

Joerger, M., Huitema, A.D.R., Boot, H., Cats, A., Doodeman, V.D., Smits, P.H.M., Vainchtein, L., Rosing, H., Meijerman, I., Zueger, M., Meulendijks, D., Cerny, T.D., Beijnen, J.H., Schellens, J.H.M., 2015. Germline TYMS genotype is highly predictive in patients with metastatic gastrointestinal malignancies receiving capecitabine-

based chemotherapy. Cancer Chemother Pharmacol 75, 763–772. https://doi.org/10.1007/S00280-015-2698-7

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., de Bakker, P.I.W., 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24, 2938. https://doi.org/10.1093/BIOINFORMATICS/BTN564

Kabat, G.C., Kim, M.Y., Manson, J.A.E., Lessin, L., Lin, J., Wassertheil-Smoller, S., Rohan, T.E., 2017. White Blood Cell Count and Total and Cause-Specific Mortality in the Women's Health Initiative. Am J Epidemiol 186, 63. https://doi.org/10.1093/AJE/KWW226

Kaburagi, T., Kizuka, Y., Kitazume, S., Taniguchi, N., 2017. The inhibitory role of α2,6-sialylation in adipogenesis. J Biol Chem 292, 2278–2286. https://doi.org/10.1074/jbc.M116.747667

Kaler, A.S., Purcell, L.C., 2019. Estimation of a significance threshold for genome-wide association studies. BMC Genomics 20, 1–8. https://doi.org/10.1186/S12864-019-5992-7

Karas, S., Innocenti, F., 2022. All You Need to Know About UGT1A1 Genetic Testing for Patients Treated With Irinotecan: A Practitioner-Friendly Guide. JCO Oncol Pract 18, 270–277. https://doi.org/10.1200/OP.21.00624

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, 1037–1048. https://doi.org/10.1177/0272989X15598529

Kassambara, A., Kosinski, M., Biecek, P., 2021. survminer: Drawing Survival Curves using 'ggplot2'. https://CRAN.R-project.org/package=survminer

Keely, S.J., Barrett, K.E., 2022. Intestinal secretory mechanisms and diarrhea. Am J Physiol Gastrointest Liver Physiol 322, G405–G420. https://doi.org/10.1152/AJPGI.00316.2021

Kerr, D.J., Gray, R., Mcconkey, C., Barnwell, J., 2000. Adjuvant chemotherapy with 5-fluorouracil, L-folinic acid and levamisole for patients with colorectal cancer: Non-randomised comparison of weekly versus four-weekly schedules — less pain, same gain. Annals of Oncology 11, 947–955. https://doi.org/10.1023/A:1008303229469

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. https://doi.org/10.1016/S1470-2045(16)30172-3

Kichaev, G., Yang, W.-Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., Pasaniuc, B., 2014. Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies. PLoS Genet 10, e1004722.

Kim, K.H., Shcheynikov, N., Wang, Y., Muallem, S., 2005. SLC26A7 is a Cl- channel regulated by intracellular pH. J Biol Chem 280, 6463–6470. https://doi.org/10.1074/jbc.M409162200

Kim, T., Wei, P., 2016. Incorporating ENCODE information into association analysis of whole genome sequencing data. BMC Proc 10, 257–261. https://doi.org/10.1186/S12919-016-0040-Y

Kim, Y., Quach, A., Das, S., Barrett, K.E., 2020. Potentiation of calcium-activated chloride secretion and barrier dysfunction may underlie EGF receptor tyrosine kinase inhibitor-induced diarrhea. Physiol Rep 8, e14490. https://doi.org/10.14814/PHY2.14490

Kindler, H.L., Friberg, G., Singh, D.A., Locker, G., Nattam, S., Kozloff, M., Taber, D.A., Karrison, T., Dachman, A., Stadler, W.M., Vokes, E.E., 2005. Phase II trial of bevacizumab plus gemcitabine in patients with advanced pancreatic cancer. J Clin Oncol 23, 8033–8040. https://doi.org/10.1200/JCO.2005.01.9661

Kircher, M., Witten, D.M., Jain, P., O'roak, B.J., Cooper, G.M., Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014 46:3 46, 310–315. https://doi.org/10.1038/ng.2892

Klastersky, J., de Naurois, J., Rolston, K., Rapoport, B., Maschmeyer, G., Aapro, M., Herrstedt, J., on behalf of the ESMO Guidelines Committee, 2016. Management of febrile neutropaenia: ESMO Clinical Practice Guidelines. Ann Oncol 27, v111–v118. https://doi.org/10.1093/ANNONC/MDW325

Kleibl, Z., Fidlerova, J., Kleiblova, P., Kormunda, S., Bilek, M., Bouskova, K., Sevcik, J., Novotny, J., 2009. Influence of dihydropyrimidine dehydrogenase gene (DPYD) coding sequence variants on the development of fluoropyrimidine-related toxicity in patients with high-grade toxicity and patients with excellent tolerance of fluoropyrimidine-based chemotherapy. Neoplasma 56, 303–316. https://doi.org/10.4149/NEO_2009_04_303

Koedoot, C.G., de Haan, R.J., Stiggelbout, A.M., Stalmeier, P.F.M., de Graeff, A., Bakker, P.J.M., de Haes, J.C.J.M., 2003. Palliative chemotherapy or best supportive care? A prospective study explaining patients' treatment preference and choice. Br J Cancer 89, 2219–2226. https://doi.org/10.1038/sj.bjc.6601445

König, I.R., 2011. Validation in Genetic Association Studies. Brief Bioinform 12, 253–258. https://doi.org/10.1093/BIB/BBQ074

Kooner, J.S., Saleheen, D., Sim, X., Sehmi, J., Zhang, W., Frossard, P., Been, L.F., Chia, K.S., Dimas, A.S., Hassanali, N., Jafar, T., Jowett, J.B.M., Li, X., Radha, V., Rees, S.D., Takeuchi, F., Young, R., Aung, T., Basit, A., Chidambaram, M., Das, D., Grundberg, E., Hedman, Å.K., Hydrie, Z.I., Islam, M., Khor, C.C., Kowlessur, S., Kristensen, M.M., Liju, S., Lim, W.Y., Matthews, D.R., Liu, J., Morris, A.P., Nica, A.C., Pinidiyapathirage, J.M., Prokopenko, I., Rasheed, A., Samuel, M., Shah, N., Shera, A.S., Small, K.S., Suo, C., Wickremasinghe, A.R., Wong, T.Y., Yang, M., Zhang, F., Abecasis, G.R., Barnett, A.H., Caulfield, M., Deloukas, P., Frayling, T.M., Froguel, P., Kato, N., Katulanda, P., Kelly, M.A., Liang, J., Mohan, V., Sanghera, D.K., Scott, J., Seielstad, M., Zimmet, P.Z., Elliott, P., Teo, Y.Y., McCarthy, M.I., Danesh, J., Tai, E.S., Chambers, J.C., 2011. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet 43, 984–989. https://doi.org/10.1038/ng.921

Krick, S., Helton, E.S., Easter, M., Bollenbecker, S., Denson, R., Zaharias, R., Cochran, P., Vang, S., Harris, E., Wells, J.M., Barnes, J.W., 2021. ST6GAL1 and α2-6 Sialylation Regulates IL-6 Expression and Secretion in Chronic Obstructive Pulmonary Disease. Front Immunol 12, 1. https://doi.org/10.3389/FIMMU.2021.693149

Kristensen, M.H., Pedersen, P.L., Melsen, G. v., Ellehauge, J., Mejer, J., 2010. Variants in the dihydropyrimidine dehydrogenase, methylenetetrahydrofolate reductase and thymidylate synthase genes predict early toxicity of 5-fluorouracil in colorectal cancer patients. J Int Med Res 38, 870–883. https://doi.org/10.1177/147323001003800313

Kroser, J.A., Metz, D.C., 1996. Evaluation of the adult patient with diarrhea. Prim Care 23, 629–647. https://doi.org/10.1016/S0095-4543(05)70352-3

Kubota, N., Suyama, M., 2022. Mapping of promoter usage QTL using RNA-seq data reveals their contributions to complex traits. PLoS Comput Biol 18. https://doi.org/10.1371/JOURNAL.PCBI.1010436

Kuehr, T., Ruff, P., Rapoport, B.L., Falk, S., Daniel, F., Jacobs, C., Davidson, N., Thaler, J., Boussard, B., Carmichael, J., 2004. Phase I/II study of first-line irinotecan combined with 5-fluorouracil and folinic acid Mayo Clinic schedule in patients with advanced colorectal cancer. BMC Cancer 4, 1–10. https://doi.org/10.1186/1471-2407-4-36/FIGURES/2

Kuipers, E.J., Grady, W.M., Lieberman, D., Seufferlein, T., Sung, J.J., Boelens, P.G., van de Velde, C.J.H., Watanabe, T., 2015. Colorectal cancer. Nat Rev Dis Primers 2015 1:1 1, 1–25. https://doi.org/10.1038/nrdp.2015.65

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. https://doi.org/10.1111/J.1365-2125.2006.02713.X

Kwakman, J.J.M., Elshot, Y.S., Punt, C.J.A., Koopman, M., 2020. Management of cytotoxic chemotherapy-induced hand-foot syndrome. Oncol Rev. https://doi.org/10.4081/ONCOL.2020.442

Lacey, S., Chung, J.Y., Lin, H., 2014. A comparison of whole genome sequencing with exome sequencing for family-based association studies. BMC Proc 8, 1–6. https://doi.org/10.1186/1753-6561-8-S1-S38/FIGURES/2

Lacouture, M.E., Anadkat, M., Jatoi, A., Garawin, T., Bohac, C., Mitchell, E., 2018. Dermatologic Toxicity Occurring During Anti-EGFR Monoclonal Inhibitor Therapy in Patients With Metastatic Colorectal Cancer: A Systematic Review. Clin Colorectal Cancer 17, 85. https://doi.org/10.1016/J.CLCC.2017.12.004

Lambrechts, D., Moisse, M., Delmar, P., Miles, D.W., Leighl, N., Escudier, B., van Cutsem, E., Bansal, A.T., Carmeliet, P., Scherer, S.J., de Haas, S., Pallaud, C., 2014. Genetic markers of bevacizumab-induced hypertension. Angiogenesis 17, 685–694. https://doi.org/10.1007/S10456-014-9424-7

Lander, E., Kruglyak, L., 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11, 241–247. https://doi.org/10.1038/NG1195-241

Largillier, R., Etienne-Grimaldi, M.C., Formento, J.L., Ciccolini, J., Nebbia, J.F., Ginot, A., Francoual, M., Renée, N., Ferrero, J.M., Foa, C., Namer, M., Lacarelle, B., Milano, G., 2006. Pharmacogenetics of capecitabine in advanced breast cancer patients. Clin Cancer Res 12, 5496–5502. https://doi.org/10.1158/1078-0432.CCR-06-0320

Law-Ping-Man, S., Toutain, F., Rieux-Laucat, F., Picard, C., Kammerer-Jacquet, S., Magérus-Chatinet, A., Dupuy, A., Adamski, H., 2018. Chronic granulomatous skin

lesions leading to a diagnosis of TAP1 deficiency syndrome. Pediatr Dermatol 35, e375–e377. https://doi.org/10.1111/PDE.13676

Lecomte, T., Ferraz, J.M., Zinzindohoué, F., Loriot, M.A., Tregouet, D.A., Landi, B., Berger, A., Cugnenc, P.H., Jian, R., Beaune, P., Laurent-Puig, P., 2004. Thymidylate synthase gene polymorphism predicts toxicity in colorectal cancer patients receiving 5-fluorouracil-based chemotherapy. Clin Cancer Res 10, 5880–5888. https://doi.org/10.1158/1078-0432.CCR-04-0169

Lecomte, T., Landi, B., Beaune, P., Laurent-Puig, P., Loriot, M.A., 2006. Glutathione S-transferase P1 polymorphism (Ile105Val) predicts cumulative neuropathy in patients receiving oxaliplatin-based chemotherapy. Clin Cancer Res 12, 3050–3056. https://doi.org/10.1158/1078-0432.CCR-05-2076

Lee, A.M., Shi, Q., Pavey, E., Alberts, S.R., Sargent, D.J., Sinicrope, F.A., Berenberg, J.L., Goldberg, R.M., Diasio, R.B., 2014. DPYD variants as predictors of 5-fluorouracil toxicity in adjuvant colon cancer treatment (NCCTG N0147). J Natl Cancer Inst 106. https://doi.org/10.1093/JNCI/DJU298

Lee, M., Lee, H.J., Bae, S., Lee, Y.S., 2008. Protein Sialylation by Sialyltransferase Involves Radiation Resistance. Molecular Cancer Research 6, 1316–1325. https://doi.org/10.1158/1541-7786.MCR-07-2209

Lee, S.H., Wray, N.R., Goddard, M.E., Visscher, P.M., 2011. Estimating Missing Heritability for Disease from Genome-wide Association Studies. Am J Hum Genet 88, 294. https://doi.org/10.1016/J.AJHG.2011.02.002

Leem, S., Jeong, H.H., Lee, J., Wee, K., Sohn, K.A., 2014. Fast detection of high-order epistatic interactions in genome-wide association studies using information theoretic measure. Comput Biol Chem 50, 19–28. https://doi.org/10.1016/J.COMPBIOLCHEM.2014.01.005

Leichman, C.G., Fleming, T.R., Muggia, F.M., Tangen, C.M., Ardalan, B., Doroshow, J.H., Meyers, F.J., Holcombe, R.F., Weiss, G.R., Mangalik, A., Macdonald, J.S., 1995. Phase II study of fluorouracil and its modulation in advanced colorectal cancer: a Southwest Oncology Group study. J Clin Oncol 13, 1303–1311. https://doi.org/10.1200/JCO.1995.13.6.1303

Lettre, G., 2014. Rare and low-frequency variants in human common diseases and other complex traits. J Med Genet 51, 705–714. https://doi.org/10.1136/JMEDGENET-2014-102437

Li, K.C., Palotie, A., Yuan, S., Bronnikov, D., Chen, D., Wei, X., Choi, O.W., Saarela, J., Peltonen, L., 2007. Finding disease candidate genes by liquid association. Genome Biol 8. https://doi.org/10.1186/GB-2007-8-10-R205

Li, M., Foo, J.N., Wang, J.Q., Low, H.Q., Tang, X.Q., Toh, K.Y., Yin, P.R., Khor, C.C., Goh, Y.F., Irwan, I.D., Xu, R.C., Andiappan, A.K., Bei, J.X., Rotzschke, O., Chen, M.H., Cheng, C.Y., Sun, L.D., Jiang, G.R., Wong, T.Y., Lin, H.L., Aung, T., Liao, Y.H., Saw, S.M., Ye, K., Ebstein, R.P., Chen, Q.K., Shi, W., Chew, S.H., Chen, J., Zhang, F.R., Li, S.P., Xu, G., Tai, E.S., Wang, L., Chen, N., Zhang, X.J., Zeng, Y.X., Zhang, H., Liu, Z.H., Yu, X.Q., Liu, J.J., 2015. Identification of new susceptibility loci for IgA nephropathy in Han Chinese. Nat Commun 6, 1–9. https://doi.org/10.1038/ncomms8270

Li, M., Kroetz, D.L., 2018. Bevacizumab-induced hypertension: Clinical presentation and molecular understanding. Pharmacol Ther 182, 152–160. https://doi.org/10.1016/J.PHARMTHERA.2017.08.012

Li, M., Mulkey, F., Jiang, C., O'Neil, B.H., Schneider, B.P., Shen, F., Friedman, P.N., Momozawa, Y., Kubo, M., Niedzwiecki, D., Hochster, H.S., Lenz, H.J., Atkins, J.N., Rugo, H.S., Halabi, S., Kelly, W.K., McLeod, H.L., Innocenti, F., Ratain, M.J., Venook, A.P., Owzar, K., Kroetz, D.L., 2018. Identification of a Genomic Region Between SLC29A1 and HSP90AB1 Associated with Risk of Bevacizumab-Induced Hypertension: CALGB 80405 (Alliance). Clin Cancer Res 24, 4734. https://doi.org/10.1158/1078-0432.CCR-17-1523

Liang, P.S., Chen, T.Y., Giovannucci, E., 2009. Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. Int J Cancer 124, 2406–2415. https://doi.org/10.1002/IJC.24191

Lin, G.L., McGinley, J.P., Drysdale, S.B., Pollard, A.J., 2018. Epidemiology and Immune Pathogenesis of Viral Sepsis. Front Immunol 9, 2147. https://doi.org/10.3389/FIMMU.2018.02147

Lin, P.I., Vance, J.M., Pericak-Vance, M.A., Martin, E.R., 2007. No gene is an island: The flip-flop phenomenon. Am J Hum Genet 80, 531–538. https://doi.org/10.1086/512133

Lin, S., Yue, J., Guan, X., Yuan, P., Wang, J., Luo, Y., Fan, Y., Cai, R., Li, Qiao, Chen, S., Zhang, P., Li, Qing, Ma, F., Xu, B., 2019. Polymorphisms of MTHFR and TYMS predict capecitabine-induced hand-foot syndrome in patients with metastatic breast cancer. Cancer Commun 39. https://doi.org/10.1186/S40880-019-0399-Z

Little, J., Higgins, J.P.T., Ioannidis, J.P.A., Moher, D., Gagnon, F., von Elm, E., Khoury, M.J., Cohen, B., Davey-Smith, G., Grimshaw, J., Scheet, P., Gwinn, M., Williamson, R.E., Zou, G.Y., Hutchings, K., Johnson, C.Y., Tait, V., Wiens, M., Golding, J., van Duijn, C., McLaughlin, J., Paterson, A., Wells, G., Fortier, I., Freedman, M., Zecevic, M., King, R., Infante-Rivard, C., Stewart, A., Birkett, N., 2009. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med 6, e22. https://doi.org/10.1371/journal.pmed.1000022

Liu, Y.J., Papasian, C.J., Liu, J.F., Hamilton, J., Deng, H.W., 2008. Is Replication the Gold Standard for Validating Genome-Wide Association Findings? PLoS One 3, e4037. https://doi.org/10.1371/JOURNAL.PONE.0004037

Loganayagam, A., Arenas Hernandez, M., Corrigan, A., Fairbanks, L., Lewis, C.M., Harper, P., Maisey, N., Ross, P., Sanderson, J.D., Marinaki, A.M., 2013. Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. Br J Cancer 108, 2505. https://doi.org/10.1038/BJC.2013.262

Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S., Hirschhorn, J.N., 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33, 177–182. https://doi.org/10.1038/NG1071

Longley, D.B., Harkin, D.P., Johnston, P.G., 2003. 5-fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer 3, 330–338. https://doi.org/10.1038/NRC1074

Ly, R.C., Schmidt, R.E., Kiel, P.J., Pratt, V.M., Schneider, B.P., Radovich, M., Offer, S.M., Diasio, R.B., Skaar, T.C., 2020. Severe Capecitabine Toxicity Associated With a Rare DPYD Variant Identified Through Whole-Genome Sequencing. JCO Precis Oncol 4, 632–638. https://doi.org/10.1200/PO.20.00067

Macaron, C., Leach, B.H., Burke, C.A., 2015. Hereditary colorectal cancer syndromes and genetic testing. J Surg Oncol 111, 103–111. https://doi.org/10.1002/JSO.23706

Madi, A., Fisher, D., Maughan, T.S., Colley, J.P., Meade, A.M., Maynard, J., Humphreys, V., Wasan, H., Adams, R.A., Idziaszczyk, S., Harris, R., Kaplan, R.S., Cheadle, J.P., 2018. Pharmacogenetic analyses of 2183 patients with advanced colorectal cancer; potential role for common dihydropyrimidine dehydrogenase variants in toxicity to chemotherapy. Eur J Cancer 102, 31–39. https://doi.org/10.1016/j.ejca.2018.07.009

Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., Cook, J.P., Schmidt, E.M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M.H., Prins, B.P., Guo, X., Bielak, L.F., Below, J.E., Bowden, D.W., Chambers, J.C., Kim, Y.J., Ng, M.C.Y., Petty, L.E., Sim, X., Zhang, W., Bennett, A.J., Bork-Jensen, J., Brummett, C.M., Canouil, M., Ec kardt, K.U., Fischer, K., Kardia, S.L.R., Kronenberg, F., Läll, K., Liu, C.T., Locke, A.E., Luan, J., Ntalla, I., Nylander, V., Schönherr, S., Schurmann, C., Yengo, L., Bottinger, E.P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J.C., Ford, I., Franco, O.H., Frayling, T.M., Giedraitis, V., Hackinger, S., Hattersley, A.T., Herder, C., Ikram, M.A., Ingelsson, M., Jørgensen, M.E., Jørgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren, C.M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K.L., Morris, A.D., Nadkarni, G., Pankow, J.S., Peters, A., Sattar, N., Stančáková, A., Strauch, K., Taylor, K.D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D.R., Dupuis, J., Peyser, P.A., Zeggini, E., Loos, R.J.F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F.S., Jukema, J.W., Palmer, C.N.A., Grallert, H., Metspalu, A., Dehghan, A., Köttgen, A., Abecasis, G.R., Meigs, J.B., Rotter, J.I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N.J., Stefansson, K., Gloyn, A.L., Morris, A.P., Boehnke, M., McCarthy, M.I., 2018. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50, 1505–1513. https://doi.org/10.1038/s41588-018-0241-6

Maher, B.S., Reimers, M.A., Riley, B.P., Kendler, K.S., 2010. Allelic heterogeneity in genetic association meta-analysis: An application to DTNBP1 and schizophrenia. Hum Hered 69, 71–79. https://doi.org/10.1159/000264445

Manchia, M., Cullis, J., Turecki, G., Rouleau, G.A., Uher, R., Alda, M., 2013. The Impact of Phenotypic and Genetic Heterogeneity on Results of Genome Wide Association Studies of Complex Diseases. PLoS One 8, 76295. https://doi.org/10.1371/JOURNAL.PONE.0076295

Mandelblatt, J.S., Huang, K., Makgoeng, S.B., Luta, G., Song, J.X., Tallarico, M., Roh, J.M., Munneke, J.R., Houlston, C.A., McGuckin, M.E., Cai, L., Hillyer, G.C., Hershman, D.L., Neugut, A.I., Isaacs, C., Kushi, L., 2015. Preliminary development and evaluation of an algorithm to identify breast cancer chemotherapy toxicities

using electronic medical records and administrative data. J Oncol Pract 11, e1–e8. https://doi.org/10.1200/JOP.2013.001288

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., MacKay, T.F.C., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753. https://doi.org/10.1038/NATURE08494

Maranville, J.C., and Cox, N.J., 2016. Pharmacogenomic variants have larger effect sizes than genetic variants associated with other dichotomous complex traits. Pharmacogenomics J, 16(4), 388–392. https://doi.org/10.1038/tpj.2015.47

Maroun, J.A., Anthony, L.B., Blais, N., Burkes, R., Dowden, S.D., Dranitsaris, G., Samson, B., Shah, A., Thirlwell, M.P., Vincent, M.D., Wong, R., 2007. Prevention and management of chemotherapy-induced diarrhea in patients with colorectal cancer: a consensus statement by the Canadian Working Group on Chemotherapy-Induced Diarrhea. Current Oncology 14, 13. https://doi.org/10.3747/CO.2007.96

Martinelli, E., Morgillo, F., Troiani, T., Tortora, G., Ciardiello, F., 2007. Panitumumab: the evidence of its therapeutic potential in metastatic colorectal cancer care. Core Evid 2, 81. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3012427/

Martínez, M.E., Jacobs, E.T., Ashbeck, E.I., Sinha, R., Lance, P., Alberts, D.S., Thompson, P.A., 2007. Meat intake, preparation methods, mutagens and colorectal adenoma recurrence. Carcinogenesis 28, 2019–2027. https://doi.org/10.1093/CARCIN/BGM179

Martinez-Balibrea, E., Abad, A., Martínez-Cards, A., Ginés, A., Valladares, M., Navarro, M., Aranda, E., Marcuello, E., Benavides, M., Massutí, B., Carrato, A., Layos, L., Manzano, J.L., Moreno, V., 2010. UGT1A and TYMS genetic variants predict toxicity and response of colorectal cancer patients treated with first-line irinotecan and fluorouracil combination therapy. Br J Cancer 103, 581–589. https://doi.org/10.1038/SJ.BJC.6605776

Mattia, E. de, Roncato, R., Fratte, C.D., Ecca, F., Toffoli, G., Cecchin, E., 2019. The use of pharmacogenetics to increase the safety of colorectal cancer patients treated with fluoropyrimidines. Cancer Drug Resistance 2, 116–130. https://doi.org/10.20517/CDR.2019.04

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, 2103–2114. https://doi.org/10.1016/S0140-6736(11)60613-2

May-Zhang, A.A., Tycksen, E., Southard-Smith, A.N., Deal, K.K., Benthal, J.T., Buehler, D.P., Adam, M., Simmons, A.J., Monaghan, J.R., Matlock, B.K., Flaherty, D.K., Potter, S.S., Lau, K.S., Southard-Smith, E.M., 2021. Combinatorial Transcriptional Profiling of Mouse and Human Enteric Neurons Identifies Shared and Disparate Subtypes In Situ. Gastroenterology 160, 755-770.e26. https://doi.org/10.1053/J.GASTRO.2020.09.032

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A., Hirschhorn, J.N., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9, 356–369. https://doi.org/10.1038/NRG2344

McCaw, Z.R., Colthurst, T., Yun, T., Furlotte, N.A., Carroll, A., Alipanahi, B., McLean, C.Y., Hormozdiari, F., 2022. DeepNull models non-linear covariate effects to improve phenotypic prediction and association power. Nat Commun 13. https://doi.org/10.1038/S41467-021-27930-0

McLeod, H.L., Sargent, D.J., Marsh, S., Green, E.M., King, C.R., Fuchs, C.S., Ramanathan, R.K., Williamson, S.K., Findlay, B.P., Thibodeau, S.N., Grothey, A., Morton, R.F., Goldberg, R.M., 2010. Pharmacogenetic predictors of adverse events and response to chemotherapy in metastatic colorectal cancer: Results from North American Gastrointestinal Intergroup Trial N9741. J Clin Oncol 28, 3227–3233. https://doi.org/10.1200/JCO.2009.21.7943

McQuade, R.M., Stojanovska, V., Abalo, R., Bornstein, J.C., Nurgali, K., 2016. Chemotherapy-Induced Constipation and Diarrhea: Pathophysiology, Current and Emerging Treatments. Front Pharmacol 7. https://doi.org/10.3389/FPHAR.2016.00414

Mersha, T.B., Martin, L.J., Biagini Myers, J.M., Kovacic, M.B., He, H., Lindsey, M., Sivaprasad, U., Chen, W., Khurana Hershey, G.K., 2015. Genomic architecture of

asthma differs by sex. Genomics 106, 15–22. https://doi.org/10.1016/j.ygeno.2015.03.003

Meulendijks, D., Rozeman, E.A., Cats, A., Sikorska, K., Joerger, M., Deenen, M.J., Beijnen, J.H., Schellens, J.H.M., 2016. Pharmacogenetic variants associated with outcome in patients with advanced gastric cancer treated with fluoropyrimidine and platinum-based triplet combinations: a pooled analysis of three prospective studies. Pharmacogenomics 2017 17:5 17, 441–451. https://doi.org/10.1038/tpj.2016.81

Meyerhardt, J.A., Tepper, J.E., Niedzwiecki, D., Hollis, D.R., McCollum, A.D., Brady, D., O'Connell, M.J., Mayer, R.J., Cummings, B., Willett, C., Macdonald, J.S., Benson, A.B., Fuchs, C.S., 2004. Impact of body mass index on outcomes and treatment-related toxicity in patients with stage II and III rectal cancer: findings from Intergroup Trial 0114. J Clin Oncol 22, 648–657. https://doi.org/10.1200/JCO.2004.07.121

Mieth, B., Kloft, M., Rodríguez, J.A., Sonnenburg, S., Vobruba, R., Morcillo-Suárez, C., Farré, X., Marigorta, U.M., Fehr, E., Dickhaus, T., Blanchard, G., Schunk, D., Navarro, A., Müller, K.R., 2016. Combining Multiple Hypothesis Testing with Machine Learning Increases the Statistical Power of Genome-wide Association Studies. Sci Rep 6. https://doi.org/10.1038/SREP36671

Mikolajewicz, N., Komarova, S. V., 2019. Meta-analytic methodology for basic research: A practical guide. Front Physiol 10, 203. https://doi.org/10.3389/FPHYS.2019.00203

Milano, G., Etienne-Grimaldi, M.-C., Mari, M., Lassalle, S., Formento, J.-L., Francoual, M., Lacour, J.-P., Hofman, P., 2008. Candidate mechanisms for capecitabine-related hand–foot syndrome. Br J Clin Pharmacol 66, 88–95. https://doi.org/10.1111/J.1365-2125.2008.03159.X

Millard, L.A.C., Davies, N.M., Gaunt, T.R., Smith, G.D., Tilling, K., 2018. Software application profile: PHESANT: A tool for performing automated phenome scans in UK Biobank. Int J Epidemiol 47, 29–35. https://doi.org/10.1093/ije/dyx204

Miura, K., Kinouchi, M., Ishida, K., Fujibuchi, W., Naitoh, T., Ogawa, H., Ando, T., Yazaki, N., Watanabe, K., Haneda, S., Shibata, C., Sasaki, I., 2010. 5FU Metabolism in Cancer and Orally-Administrable 5FU Drugs. Cancers (Basel) 2, 1717. https://doi.org/10.3390/CANCERS2031717

Modest, D.P., Karthaus, M., Fruehauf, S., Graeven, U., Müller, L., König, A.O., von Weikersthal, L.F., Caca, K., Kretzschmar, A., Goekkurt, E., Haas, S., Kurreck, A., Stahler, A., Held, S., Jarosch, A., Horst, D., Reinacher-Schick, A., Kasper, S., Heinemann, V., Stintzing, S., Trarbach, T., 2022. Panitumumab Plus Fluorouracil

and Folinic Acid Versus Fluorouracil and Folinic Acid Alone as Maintenance Therapy in RAS Wild-Type Metastatic Colorectal Cancer: The Randomized PANAMA Trial (AIO KRK 0212). J Clin Oncol 40, 72–82. https://doi.org/10.1200/JCO.21.01332

Modest, D.P., Pant, S., Sartore-Bianchi, A., 2019. Treatment sequencing in metastatic colorectal cancer. Eur J Cancer 109, 70–83. https://doi.org/10.1016/J.EJCA.2018.12.019

Monahan, K.J., Bradshaw, N., Dolwani, S., Desouza, B., Dunlop, M.G., East, J.E., Ilyas, M., Kaur, A., Lalloo, F., Latchford, A., Rutter, M.D., Tomlinson, I., Thomas, H.J.W., Hill, J., 2020. Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG). Gut 69, 411–444. https://doi.org/10.1136/GUTJNL-2019-319915

Mongiovi, J.M., Zirpoli, G.R., Cannioto, R., Sucheston-Campbell, L.E., Hershman, D.L., Unger, J.M., Moore, H.C.F., Stewart, J.A., Isaacs, C., Hobday, T.J., Salim, M., Hortobagyi, G.N., Gralow, J.R., Thomas Budd, G., Albain, K.S., Ambrosone, C.B., McCann, S.E., 2018. Associations between self-reported diet during treatment and chemotherapy-induced peripheral neuropathy in a cooperative group trial (S0221). Breast Cancer Research 20. https://doi.org/10.1186/s13058-018-1077-9

Montagut, C., Dalmases, A., Bellosillo, B., Crespo, M., Pairet, S., Iglesias, M., Salido, M., Gallen, M., Marsters, S., Tsai, S.P., Minoche, A., Somasekar, S., Serrano, S., Himmelbauer, H., Bellmunt, J., Rovira, A., Settleman, J., Bosch, F., Albanell, J., 2012. Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. Nat Med 2012 18:2 18, 221–223. https://doi.org/10.1038/nm.2609

Moore, C., Jacobson, S., 2021. genpwr: Power Calculations Under Genetic Model Misspecification. https://CRAN.R-project.org/package=genpwr

Moran, R.G., Keyomarsi, K., 1987. Biochemical rationale for the synergism of 5-fluorouracil and folinic acid. NCI Monogr 159–163. https://pubmed.ncbi.nlm.nih.gov/2963229/

Morawska, K., Goirand, F., Marceau, L., Devaux, M., Cueff, A., Bertaut, A., Vincent, J., Bengrine-Lefevre, L., Ghiringhelli, F., Schmitt, A., 2018. 5FU therapeutic drug monitoring as a valuable option to reduce toxicity in patients with gastrointestinal cancer. Oncotarget 9, 11559. https://doi.org/10.18632/ONCOTARGET.24338

Morel, A., Boisdron-Celle, M., Fey, L., Soulie, P., Craipeau, M.C., Traore, S., Gamelin, E., 2006. Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. Mol Cancer Ther 5, 2895–2904. https://doi.org/10.1158/1535-7163.MCT-06-0327

Müller, K., Wickham, H., 2021. tibble: Simple Data Frames. https://CRAN.R-project.org/package=tibble

Nakaoka, H., Inoue, I., 2009. Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. Hum Genet 2009 54:11 54, 615–623. https://doi.org/10.1038/jhg.2009.95

Nasirikenari, M., Chandrasekaran, E. v., Matta, K.L., Segal, B.H., Bogner, P.N., Lugade, A.A., Thanavala, Y., Lee, J.J., Lau, J.T.Y., 2010. Altered eosinophil profile in mice with ST6Gal-1 deficiency: an additional role for ST6Gal-1 generated by the P1 promoter in regulating allergic inflammation. J Leukoc Biol 87, 457–466. https://doi.org/10.1189/jlb.1108704

Nasirikenari, M., Lugade, A.A., Neelamegham, S., Gao, Z., Moremen, K.W., Bogner, P.N., Thanavala, Y., Lau, J.T.Y., 2019. Recombinant Sialyltransferase Infusion Mitigates Infection-Driven Acute Lung Inflammation. Front Immunol 10, 48. https://doi.org/10.3389/fimmu.2019.00048

Nasirikenari, M., Segal, B.H., Ostberg, J.R., Urbasic, A., Lau, J.T.Y., 2006. Altered granulopoietic profile and exaggerated acute neutrophilic inflammation in mice with targeted deficiency in the sialyltransferase ST6Gal I. Blood 108, 3397–3405. https://doi.org/10.1182/blood-2006-04-014779

National Cancer Institute, 2017. Common terminology criteria for adverse events. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm#ctc_60

Nayak, M.G., George, A., Vidyasagar, M., Mathew, S., Nayak, S., Nayak, B.S., Shashidhara, Y., Kamath, A., 2017. Quality of Life among Cancer Patients. Indian J Palliat Care 23, 445. https://doi.org/10.4103/IJPC.IJPC_82_17

Neugut, A.I., Prigerson, H.G., 2017. Curative, Life-Extending, and Palliative Chemotherapy: New Outcomes Need New Names. Oncologist 22, 883. https://doi.org/10.1634/THEONCOLOGIST.2017-0041

Neuwirth, E., 2014. RColorBrewer: ColorBrewer Palettes. https://CRAN.R-project.org/package=RColorBrewer

NICE guidelines, 2012a. Bevacizumab and cetuximab for the treatment of metastatic colorectal cancer. Technology appraisal guidance [TA118]. https://www.nice.org.uk/guidance/ta118

NICE guidelines, 2012b. Neutropenic sepsis: prevention and management in people with cancer. Clinical guideline 151. https://www.nice.org.uk/guidance/cg151

NICE guidelines, 2017. Cetuximab and panitumumab for previously untreated metastatic colorectal cancer. Technology appraisal guidance [TA439]. https://www.nice.org.uk/guidance/ta439

Nicholls, H.L., John, C.R., Watson, D.S., Munroe, P.B., Barnes, M.R., Cabrera, C.P., 2020. Reaching the End-Game for GWAS: Machine Learning Approaches for the Prioritization of Complex Disease Loci. Front Genet 11. https://doi.org/10.3389/FGENE.2020.00350

O'Donnell, P.H., Trubetskoy, V., Nurhussein-Patterson, A., Hall, J.P., Nath, A., Huo, D., Fleming, G.F., Ingle, J.N., Abramson, V.G., Morrow, P.K., Storniolo, A.M., Forero, A., van Poznak, C., Liu, M.C., Chang, J.C., Merkel, D.E., Peppercorn, J.M., Rugo, H.S., Dees, E.C., Hahn, O.M., Hoffman, P.C., Rosner, G.L., Huang, R.S., Ratain, M.J., Cox, N., Olopade, O.I., Wolff, A.C., Dolan, M.E., Nanda, R., 2020. Clinical evaluation of germline polymorphisms associated with capecitabine toxicity in breast cancer: TBCRC-015. Breast Cancer Res Treat 181, 623–633. https://doi.org/10.1007/S10549-020-05603-8

Oetting, W.S., Jacobson, P.A., Israni, A.K., 2017. Validation is Critical for GWAS-based Associations. Am J Transplant 17, 318. https://doi.org/10.1111/AJT.14051

Oostendorp, L.J.M., Stalmeier, P.F., Pasker-De Jong, P.C., van der Graaf, W.T., Ottevanger, P.B., 2010. Systematic review of benefits and risks of second-line irinotecan monotherapy for advanced colorectal cancer. Anticancer Drugs 21, 749–758. https://doi.org/10.1097/CAD.0B013E32833C57CF

Österlund, P., Ruotsalainen, T., Peuhkuri, K., Korpela, R., Ollus, A., Ikonen, M., Joensuu, H., Elomaa, I., 2004. Lactose intolerance associated with adjuvant 5-fluorouracil-based chemotherapy for colorectal cancer. Clin Gastroenterol Hepatol 2, 696–703. https://doi.org/10.1016/S1542-3565(04)00293-9

Oswald, D.M., Zhou, J.Y., Jones, M.B., Cobb, B.A., 2020. Disruption of Hepatocyte Sialylation Drives a T cell-Dependent Pro-Inflammatory Immune Tone. Glycoconj J 37, 395. https://doi.org/10.1007/S10719-020-09918-Y

Ouyang, R., Li, H., Xia, J., Wang, X., Zheng, X., Huang, Y., Meng, Z., Gao, Y., Qian, Z., Liu, F., Lu, X., Shi, Y., Shang, J., Liu, J., Deng, G., Zheng, Y., Yan, H., Zhang, W., Qiao, L., Jiang, X., Wang, H., Zhong, G., Li, B., Chen, J., 2021. Lower platelet counts were associated with 90-day adverse outcomes in acute-on-chronic liver disease patients. Ann Palliat Med 10, 9342–9353. https://doi.org/10.21037/APM-21-1019

Pachman, D.R., Qin, R., Seisler, D.K., Smith, E.M.L., Beutler, A.S., Ta, L.E., Lafky, J.M., Wagner-Johnston, N.D., Ruddy, K.J., Dakhil, S., Staff, N.P., Grothey, A., Loprinzi, C.L., 2015. Clinical course of oxaliplatin-induced neuropathy: Results from the randomized phase III trial N08CB (Alliance). J Clin Oncol 33, 3416–3422. https://doi.org/10.1200/JCO.2014.58.8533

Pain, O., Dudbridge, F., Ronald, A., 2018. Are your covariates under control? How normalization can re-introduce covariate effects. Eur J Hum Genet 2018 26:8 26, 1194–1201. https://doi.org/10.1038/s41431-018-0159-6

Panagiotou, O.A., Willer, C.J., Hirschhorn, J.N., Ioannidis, J.P.A., 2013. The Power of Meta-Analysis in Genome Wide Association Studies. Annu Rev Genomics Hum Genet 14, 441. https://doi.org/10.1146/ANNUREV-GENOM-091212-153520

Panoutsopoulou, K., Tachmazidou, I., Zeggini, E., 2013. In search of low-frequency and rare variants affecting complex traits. Hum Mol Genet 22, R16. https://doi.org/10.1093/HMG/DDT376

Paré, G., Mao, S., Deng, W.Q., 2017. A machine-learning heuristic to improve gene score prediction of polygenic traits. Sci Rep 7. https://doi.org/10.1038/S41598-017-13056-1

Park, H.A., Seibold, P., Edelmann, D., Benner, A., Canzian, F., Alwers, E., Jansen, L., Schneider, M., Hoffmeister, M., Brenner, H., Chang-Claude, J., 2022. Validation of Genetic Markers Associated with Survival in Colorectal Cancer Patients Treated with Oxaliplatin-Based Chemotherapy. Cancer Epidemiol Biomarkers Prev 31, 352–361. https://doi.org/10.1158/1055-9965.EPI-21-0814

Park, I.H., Im, S.A., Jung, K.H., Sohn, J.H., Park, Y.H., Lee, K.S., Sim, S.H., Park, K.H., Kim, J.H., Nam, B.H., Kim, H.J., Kim, T.Y., Lee, K.H., Kim, S.B., Ahn, J.H., Lee, S., Ro, J., 2019. Randomized Open Label Phase III Trial of Irinotecan Plus Capecitabine versus Capecitabine Monotherapy in Patients with Metastatic Breast Cancer Previously Treated with Anthracycline and Taxane: PROCEED Trial (KCSG BR 11-01). Cancer Research and Treatment : Official Journal of Korean Cancer Association 51, 43. https://doi.org/10.4143/CRT.2017.562

Parnes, H.L., Fung, E., Schifer, C.A., 1994. Chemotherapy-Induced Lactose Intolerance in Adults. Cancer 74, 1629–1633. https://doi.org/10.1002/1097-0142(19940901)74:5

Pearson, T.A., Manolio, T.A., 2008. How to Interpret a Genome-wide Association Study. JAMA 299, 1335–1344. https://doi.org/10.1001/JAMA.299.11.1335

Pellicer, M., García-González, X., García, M.I., Robles, L., Grávalos, C., García-Alfonso, P., Pachón, V., Longo, F., Martínez, V., Blanco, C., Iglesias, I., Sanjurjo, M., López-Fernández, L.A., 2017. Identification of new SNPs associated with severe toxicity to capecitabine. Pharmacol Res 120, 133–137. https://doi.org/10.1016/J.PHRS.2017.03.021

Peng, Z., Wang, Q., Gao, J., Ji, Z., Yuan, J., Tian, Y., Shen, L., 2013. Association between GSTP1 Ile105Val polymorphism and oxaliplatin-induced neuropathy: a systematic review and meta-analysis. Cancer Chemother Pharmacol 72, 305–314. https://doi.org/10.1007/S00280-013-2194-X

Pergolizzi, J. V., Taylor, R., LeQuang, J.A., Zampogna, G., Raffa, R.B., 2017. Concise review of the management of iatrogenic emesis using cannabinoids: Emphasis on nabilone for chemotherapy-induced nausea and vomiting. Cancer Chemother Pharmacol 79, 467–477. https://doi.org/10.1007/S00280-017-3257-1

Petrelli, F., Borgonovo, K., Barni, S., 2013. 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 8, 173–181. https://doi.org/10.1007/S11523-013-0257-X

Petrovic, S., Ju, X., Barone, S., Seidler, U., Alper, S.L., Lohi, H., Kere, J., Soleimani, M., 2003. Identification of a basolateral Cl-/HCO3- exchanger specific to gastric parietal cells. Am J Physiol Gastrointest Liver Physiol 284. https://doi.org/10.1152/AJPGI.00454.2002

Piedbois, P., Buyse, M., Rustum, Y., Machover, D., Erlichman, C., Carlson, R.W., Valone, F., Labianca, R., Doroshow, J.H., Petrelli, N., 1992. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. Advanced Colorectal Cancer Meta-Analysis Project. J Clin Oncol 10, 896–903. https://doi.org/10.1200/JCO.1992.10.6.896

Pini, A., Garella, R., Idrizaj, E., Calosi, L., Baccari, M.C., Vannucchi, M.G., 2016. Glucagon-like peptide 2 counteracts the mucosal damage and the neuropathy induced by chronic treatment with cisplatin in the mouse gastric fundus. Neurogastroenterol motil 28, 206–216. https://doi.org/10.1111/NMO.12712

Pinto, C., Barone, C.A., Girolomoni, G., Russi, E.G., Merlano, M.C., Ferrari, D., Maiello, E., 2011. Management of Skin Toxicity Associated with Cetuximab Treatment in Combination with Chemotherapy or Radiotherapy. Oncologist 16, 228. https://doi.org/10.1634/THEONCOLOGIST.2010-0298

Pinto, N., Dolan, M.E., 2011. Clinically Relevant Genetic Variations in Drug Metabolizing Enzymes. Curr Drug Metab 12, 487. https://doi.org/10.2174/138920011795495321

Price, T.J., Hardingham, J.E., Lee, C.K., Weickhardt, A., Townsend, A.R., Wrin, J.W., Chua, A., Shivasami, A., Cummins, M.M., Murone, C., Tebbutt, N.C., 2011. Impact of KRAS and BRAF Gene Mutation Status on Outcomes From the Phase III AGITG MAX Trial of Capecitabine Alone or in Combination With Bevacizumab and Mitomycin in Advanced Colorectal Cancer. J Clin Oncol 29, 2675–2682. https://doi.org/10.1200/JCO.2010.34.5520

Price, T.J., Peeters, M., Kim, T.W., Li, J., Cascinu, S., Ruff, P., Suresh, A.S., Thomas, A., Tjulandin, S., Zhang, K., Murugappan, S., Sidhu, R., 2014. Panitumumab versus cetuximab in patients with chemotherapy-refractory wild-type KRAS exon 2 metastatic colorectal cancer (ASPECCT): a randomised, multicentre, open-label, non-inferiority phase 3 study. Lancet Oncol 15, 569–579. https://doi.org/10.1016/S1470-2045(14)70118-4

Pritchard, J.K., 2001. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69, 124–137. https://doi.org/10.1086/321272

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., Frishman, D., 2010. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336. https://doi.org/10.1093/BIOINFORMATICS/BTQ419

Punch, P.R., Irons, E.E., Manhardt, C.T., Marathe, H., Lau, J.T.Y., 2020. The sialyltransferase ST6GAL1 protects against radiation-induced gastrointestinal damage. Glycobiology 30, 446–453. https://doi.org/10.1093/GLYCOB/CWZ108

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., 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, 559–575. https://doi.org/10.1086/519795

Qin, S., Li, J., Wang, L., Xu, J., Cheng, Y., Bai, Y., Li, W., Xu, N., Lin, L.Z., Wu, Q., Li, Y., Yang, J., Pan, H., Ouyang, X., Qiu, W., Wu, K., Xiong, J., Dai, G., Liang, H., Hu, C., Zhang, J., Tao, M., Yao, Q., Wang, J., Chen, J., Peter Eggleton, S., Liu, T., 2018.

Efficacy and Tolerability of First-Line Cetuximab Plus Leucovorin, Fluorouracil, and Oxaliplatin (FOLFOX-4) Versus FOLFOX-4 in Patients With RAS Wild-Type Metastatic Colorectal Cancer: The Open-Label, Randomized, Phase III TAILOR Trial. J Clin Oncol 36, 3031. https://doi.org/10.1200/JCO.2018.78.3183

Qu, W., Gurdziel, K., Pique-Regi, R., Ruden, D.M., 2017. Identification of splicing quantitative trait loci (sQTL) in Drosophila melanogaster with developmental lead (Pb2+) exposure. Front Genet 8, 145. https://doi.org/10.3389/FGENE.2017.00145/FULL

Quintanilha, J.C.F., Wang, J., Sibley, A.B., Jiang, C., Etheridge, A.S., Shen, F., Jiang, G., Mulkey, F., Patel, J.N., Hertz, D.L., Dees, E.C., McLeod, H.L., Bertagnolli, M., Rugo, H., Kindler, H.L., Kelly, W.K., Ratain, M.J., Kroetz, D.L., Owzar, K., Schneider, B.P., Lin, D., Innocenti, F., 2022. Bevacizumab-induced hypertension and proteinuria: a genome-wide study of more than 1000 patients. Br J Cancer 126, 265–274. https://doi.org/10.1038/S41416-021-01557-W

Ranpura, V., Pulipati, B., Chu, D., Zhu, X., Wu, S., 2010. Increased Risk of High-Grade Hypertension With Bevacizumab in Cancer Patients: A Meta-Analysis. Am J Hypertens 23, 460–468. https://doi.org/10.1038/AJH.2010.25

Rao, M., Gershon, M.D., 2016. The bowel and beyond: the enteric nervous system in neurological disorders. Nat Rev Gastroenterol & Hepatol 2016 13:9 13, 517–528. https://doi.org/10.1038/nrgastro.2016.107

Raufman, J.P., Collins, S.M., Pandol, S.J., Korman, L.Y., Collen, M.J., Cornelius, M.J., Feld, M.K., McCarthy, D.M., Gardner, J.D., Jensen, R.T., 1983. Reliability of Symptoms in Assessing Control of Gastric Acid Secretion in Patients With Zollinger-Ellison Syndrome. Gastroenterology 84, 108–113. https://doi.org/10.1016/S0016-5085(83)80173-5

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL https://www.R-project.org/.

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, 3769–3792. https://doi.org/10.1002/SIM.6605

Reich, D.E., Lander, E.S., 2001. On the allelic spectrum of human disease. Trends Genet 17, 502–510. https://doi.org/10.1016/S0168-9525(01)02410-6

Riera, P., Artigas-Baleri, A., Salazar, J., Sebio, A., Virgili, A.C., Arranz, M.J., Páez, D., 2020. ABCB1 Genetic Variants as Predictors of Irinotecan-Induced Severe

Gastrointestinal Toxicity in Metastatic Colorectal Cancer Patients. Front Pharmacol 11, 973. https://doi.org/10.3389/FPHAR.2020.00973/BIBTEX

Risch, N., Merikangas, K., 1996. The future of genetic studies of complex human diseases. Science (1979) 273, 1516–1517. https://doi.org/10.1126/SCIENCE.273.5281.1516

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. https://doi.org/10.1200/JCO.2013.51.1857

Rosmarin, D., Palles, C., Pagnamenta, A., Kaur, K., Pita, G., Martin, M., Domingo, E., Jones, A., Howarth, K., Freeman-Mills, L., Johnstone, E., Wang, H., Love, S., Scudder, C., Julier, P., Fernández-Rozadilla, C., Ruiz-Ponte, C., Carracedo, A., Castellvi-Bel, S., Castells, A., Gonzalez-Neira, A., Taylor, J., Kerr, R., Kerr, D., Tomlinson, I., 2015. A candidate gene study of capecitabine-related toxicity in colorectal cancer identifies new toxicity variants at DPYD and a putative role for ENOSF1 rather than TYMS. Gut 64, 111–120. https://doi.org/10.1136/GUTJNL-2013-306571/-/DC1

Ruzzo, A., Graziano, F., Galli, Fabio, Giacomini, E., Floriani, I., Galli, Francesca, Rulli, E., Lonardi, S., Ronzoni, M., Massidda, B., Zagonel, V., Pella, N., Mucciarini, C., Labianca, R., Ionta, M.T., Veltri, E., Sozzi, P., Barni, S., Ricci, V., Foltran, L., Nicolini, M., Biondi, E., Bramati, A., Turci, D., Lazzarelli, S., Verusio, C., Bergamo, F., Sobrero, A., Frontini, L., Magnani, M., 2014. Genetic markers for toxicity of adjuvant oxaliplatin and fluoropyrimidines in the phase III TOSCA trial in high-risk colon cancer patients. Sci Rep 4. https://doi.org/10.1038/SREP06828

Ruzzo, A., Graziano, F., Loupakis, F., Rulli, E., Canestrari, E., Santini, D., Catalano, V., Ficarelli, R., Maltese, P., Bisonni, R., Masi, G., Schiavon, G., Giordani, P., Giustini, L., Falcone, A., Tonini, G., Silva, R., Mattioli, R., Floriani, I., Magnani, M., 2007.

Pharmacogenetic profiling in patients with advanced colorectal cancer treated with first-line FOLFOX-4 chemotherapy. J Clin Oncol 25, 1247–1254. https://doi.org/10.1200/JCO.2006.08.1844

Saade, S., Cazier, J.B., Ghassibe-Sabbagh, M., Youhanna, S., Badro, D.A., Kamatani, Y., Hager, J., Yeretzian, J.S., El-Khazen, G., Haber, M., Salloum, A.K., Douaihy, B., Othman, R., Shasha, N., Kabbani, S., Bayeh, H. el, Chammas, E., Farrall, M., Gauguier, D., Platt, D.E., Zalloua, P.A., 2011. Large Scale Association Analysis Identifies Three Susceptibility Loci for Coronary Artery Disease. PLoS One 6, 29427. https://doi.org/10.1371/JOURNAL.PONE.0029427

Saha, S., Bardelli, A., Buckhaults, P., Velculescu, V.E., Rago, C., st. Croix, B., Romans, K.E., Choti, M.A., Lengauer, C., Kinzler, K.W., Vogelstein, B., 2001. A phosphatase associated with metastasis of colorectal cancer. Science (1979) 294, 1343–1346. https://doi.org/10.1126/science.1065817

Saif, M.W., Reardon, J., 2005. Management of oxaliplatin-induced peripheral neuropathy. Ther Clin Risk Manag 1, 249. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1661634/

Salgueiro, N., Veiga, I., Fragoso, M., Sousa, O., Costa, N., Pellon, M.L., Sanches, E., Guimarães Dos Santos, J.G., Teixeira, M.R., Castedo, S., 2004. Mutations in exon 14 of dihydropyrimidine dehydrogenase and 5-Fluorouracil toxicity in Portuguese colorectal cancer patients. Genet Med 6, 102–107. https://doi.org/10.1097/01.GIM.0000118061.66602.A5

Schaid, D.J., Chen, W., Larson, N.B., 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 19, 491. https://doi.org/10.1038/S41576-018-0016-Z

Schmitt, A.O., Aßmus, J., Bortfeldt, R.H., Brockmann, G.A., 2010. CandiSNPer: a web tool for the identification of candidate SNPs for causal variants. Bioinformatics 26, 969–970. https://doi.org/10.1093/BIOINFORMATICS/BTQ068

Schneider, B.P., Li, L., Shen, F., Miller, K.D., Radovich, M., O'Neill, A., Gray, R.J., Lane, D., Flockhart, D.A., Jiang, G., Wang, Z., Lai, D., Koller, D., Pratt, J.H., Dang, C.T., Northfelt, D., Perez, E.A., Shenkier, T., Cobleigh, M., Smith, M.L., Railey, E., Partridge, A., Gralow, J., Sparano, J., Davidson, N.E., Foroud, T., Sledge, G.W., 2014. Genetic variant predicts bevacizumab-induced hypertension in ECOG-5103 and ECOG-2100. Br J Cancer 111, 1241–1248. https://doi.org/10.1038/BJC.2014.430

Schork, N.J., Murray, S.S., Frazer, K.A., Topol, E.J., 2009. Common vs. Rare Allele Hypotheses for Complex Diseases. Curr Opin Genet Dev 19, 212. https://doi.org/10.1016/J.GDE.2009.04.010

Schultz, M.J., Swindall, A.F., Wright, J.W., Sztul, E.S., Landen, C.N., Bellis, S.L., 2013. ST6Gal-I sialyltransferase confers cisplatin resistance in ovarian tumor cells. J Ovarian Res 6, 25. https://doi.org/10.1186/1757-2215-6-25

Schwab, M., Zanger, U.M., Marx, C., Schaeffeler, E., Klein, K., Dippon, J., Kerb, R., Blievernicht, J., Fischer, J., Hofmann, U., Bokemeyer, C., Eichelbaum, M., 2008. Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: A prospective clinical trial by the German 5FU toxicity study group. J Clin Oncol 26, 2131–2138. https://doi.org/10.1200/JCO.2006.10.4182

Secombe, K.R., van Sebille, Y.Z.A., Mayo, B.J., Coller, J.K., Gibson, R.J., Bowen, J.M., 2020. Diarrhea Induced by Small Molecule Tyrosine Kinase Inhibitors Compared With Chemotherapy: Potential Role of the Microbiome. Integr Cancer Ther 19. https://doi.org/10.1177/1534735420928493

Segré, A. v, DIAGRAM consortium, MAGIC investigators, Groop, L., Mootha, V.K., Daly, M.J., Altshuler, D., 2010. Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. PLoS Genet 6, e1001058. https://doi.org/10.1371/JOURNAL.PGEN.1001058

Seo, S.H., Kim, S.E., Kang, Y.K., Ryoo, B.Y., Ryu, M.H., Jeong, J.H., Kang, S.S., Yang, M., Lee, J.E., Sung, M.K., 2016. Association of nutritional status-related indices and chemotherapy-induced adverse events in gastric cancer patients. BMC Cancer 16, 1–9. https://doi.org/10.1186/S12885-016-2934-5/TABLES/6

Sham, P.C., Purcell, S.M., 2014. Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet 15, 335–346. https://doi.org/10.1038/NRG3706

Shao, S., Niu, Y., Zhang, X., Kong, R., Wang, J., Liu, L., Luo, X., Zhang, J., Song, R., 2016. Opposite Associations between Individual KIAA0319 Polymorphisms and Developmental Dyslexia Risk across Populations: A Stratified Meta-Analysis by the Study Population. Sci Rep 6, 1–9. https://doi.org/10.1038/srep30454

Sharma, R., Hoskins, J.M., Rivory, L.P., Zucknick, M., London, R., Liddle, C., Clarke, S.J., 2008. Thymidylate synthase and methylenetetrahydrofolate reductase gene polymorphisms and toxicity to capecitabine in advanced colorectal cancer patients. Clin Cancer Res 14, 817–825. https://doi.org/10.1158/1078-0432.CCR-07-0425

Shastry, B.S., 2009. SNPs: impact on gene function and phenotype. Methods Mol Biol 578, 3–22. https://doi.org/10.1007/978-1-60327-411-1_1

Shih, T., Lindley, C., 2006. Bevacizumab: an angiogenesis inhibitor for the treatment of solid malignancies. Clin Ther 28, 1779–1802. https://doi.org/10.1016/J.CLINTHERA.2006.11.015

Shirley, B.C., Mucaki, E.J., Rogan, P.K., Buratti, E., Pim Pijnappel, W., 2019. Pan-cancer repository of validated natural and cryptic mRNA splicing mutations. F1000Research 2019 7:1908 7, 1908. https://doi.org/10.12688/f1000research.17204.3

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539. https://doi.org/10.1038/MSB.2011.75

Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., Ng, P.C., 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40. https://doi.org/10.1093/NAR/GKS539

Slatkin, M., 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nat Rev Gen 2008 9:6 9, 477–485. https://doi.org/10.1038/nrg2361

Slim, L., Chatelain, C., Azencott, C.A., Vert, J.P., 2020. Novel methods for epistasis detection in genome-wide association studies. PLoS One 15, e0242927. https://doi.org/10.1371/JOURNAL.PONE.0242927

Smedley, D., Smith, K.R., Martin, A., Thomas, E.A., McDonagh, E.M., Cipriani, V., Ellingford, J.M., Arno, G., Tucci, A., Vandrovcova, J., Chan, G., Williams, H.J., Ratnaike, T., Wei, W., Stirrups, K., Ibanez, K., Moutsianas, L., Wielscher, M., Need, A., Barnes, M.R., Vestito, L., Buchanan, J., Wordsworth, S., Ashford, S., Rehmström, K., Li, E., Fuller, G., Twiss, P., Spasic-Boskovic, O., Halsall, S., Floto, R.A., Poole, K., Wagner, A., Mehta, S.G., Gurnell, M., Burrows, N., James, R., Penkett, C., Dewhurst, E., Gräf, S., Mapeta, R., Kasanicki, M., Haworth, A., Savage, H., Babcock, M., Reese, M.G., Bale, M., Baple, E., Boustred, C., Brittain, H., de Burca, A., Bleda, M., Devereau, A., Halai, D., Haraldsdottir, E., Hyder, Z., Kasperaviciute, D., Patch, C., Polychronopoulos, D., Matchan, A., Sultana, R., Ryten, M., Tavares, A.L.T., Tregidgo, C., Turnbull, C., Welland, M., Wood, S., Snow, C., Williams, E., Leigh, S., Foulger, R.E., Daugherty, L.C., Niblock, O., Leong, I.U.S., Wright, C.F., Davies, J., Crichton, C., Welch, J., Woods, K., Abulhoul, L., Aurora, P.,

Bockenhauer, D., Broomfield, A., Cleary, M.A., Lam, T., Dattani, M., Footitt, E., Ganesan, V., Grunewald, S., Compeyrot-Lacassagne, S., Muntoni, F., Pilkington, C., Quinlivan, R., Thapar, N., Wallis, C., Wedderburn, L.R., Worth, A., Bueser, T., Compton, C., Deshpande, C., Fassihi, H., Haque, E., Izatt, L., Josifova, D., Mohammed, S., Robert, L., Rose, S., Ruddy, D., Sarkany, R., Say, G., Shaw, A.C., Wolejko, A., Habib, B., Burns, G., Hunter, S., Grocock, R.J., Humphray, S.J., Robinson, P.N., Haendel, M., Simpson, M.A., Banka, S., Clayton-Smith, J., Douzgou, S., Hall, G., Thomas, H.B., O'Keefe, R.T., Michaelides, M., Moore, A.T., Malka, S., Pontikos, N., Browning, A.C., Straub, V., Gorman, G.S., Horvath, R., Quinton, R., Schaefer, A.M., Yu-Wai-Man, P., Turnbull, D.M., McFarland, R., Taylor, R.W., O'Connor, E., Yip, J., Newland, K., Morris, H.R., Polke, J., Wood, N.W., Campbell, C., Camps, C., Gibson, K., Koelling, N., Lester, T., Németh, A.H., Palles, C., Patel, S., Roy, N.B.A., Sen, A., Taylor, J., Cacheiro, P., Jacobsen, J.O., Seaby, E.G., Davison, V., Chitty, L., Douglas, A., Naresh, K., McMullan, D., Ellard, S., Temple, I.K., Mumford, A.D., Wilson, G., Beales, P., Bitner-Glindzicz, M., Black, G., Bradley, J.R., Brennan, P., Burn, J., Chinnery, P.F., Elliott, P., Flinter, F., Houlden, H., Irving, M., Newman, W., Rahman, S., Sayer, J.A., Taylor, J.C., Webster, A.R., Wilkie, A.O.M., Ouwehand, W.H., Raymond, F.L., Chisholm, J., Hill, S., Bentley, D., Scott, R.H., Fowler, T., Rendon, A., Caulfield, M., 2021. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. N Engl J of Med 385, 1868–1880. https://doi.org/10.1056/NEJMoa2035790.

Soulié, P., Raymond, E., Brienza, S., Cvitkovic, E., 1997. Oxaliplatin: the first DACH platinum in clinical practice. Bull Cancer 84, 665–73. https://pubmed.ncbi.nlm.nih.gov/9295871/

Southam, L., Panoutsopoulou, K., Rayner, N.W., Chapman, K., Durrant, C., Ferreira, T., Arden, N., Carr, A., Deloukas, P., Doherty, M., Loughlin, J., McCaskie, A., Ollier, W.E.R., Ralston, S., Spector, T.D., Valdes, A.M., Wallis, G.A., Wilkinson, J.M., Marchini, J., Zeggini, E., 2011. The effect of genome-wide association scan quality control on imputation outcome for common variants. Eur J of Hum Genet 19, 610. https://doi.org/10.1038/EJHG.2010.242

Spagnoli, C., Pisani, F., di Mario, F., Leandro, G., Gaiani, F., De'angelis, G.L., Fusco, C., 2018. Peripheral neuropathy and gastroenterologic disorders: an overview on an underrecognized association. Acta biomedica 89, 22–32. https://doi.org/10.23750/ABM.V89I9-S.7956

Stein, A., Arnold, D., 2012. Oxaliplatin: a review of approved uses. Expert Opin Pharmacother 13, 125–137. https://doi.org/10.1517/14656566.2012.643870

Stein, A., Voigt, W., Jordan, K., 2010. Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management. Ther Adv Med Oncol 2, 51. https://doi.org/10.1177/1758834009355164

Stintzing, S., Fischer von Weikersthal, L., Vehling-Kaiser, U., Stauch, M., Hass, H.G., Dietzfelbinger, H., Oruzio, D., Klein, S., Zellmann, K., Decker, T., Schulze, M., Abenhardt, W., Puchtler, G., Kappauf, H., Mittermüller, J., Haberl, C., Giessen, C., Moosmann, N., Heinemann, V., 2011. Correlation of capecitabine-induced skin toxicity with treatment efficacy in patients with metastatic colorectal cancer: results from the German AIO KRK-0104 trial. Br J Cancer 2011 105:2 105, 206–211. https://doi.org/10.1038/bjc.2011.227

Suzuki, A., Guerrini, M.M., Yamamoto, K., 2021. Functional genomics of autoimmune diseases. Ann Rheum Dis 80, 689–697. https://doi.org/10.1136/annrheumdis-2019-216794

Sweet, C., Sharma, A., Lipscomb, G., 2010. Recurrent nausea, vomiting and abdominal pain due to hypothyroidism. BMJ Case Rep 2010, bcr1120092461. https://doi.org/10.1136/bcr.11.2009.2461

Taché, Y., Stephens, R.L., Ishikawa, T., 1989. Central Nervous System Action of TRH to Influence Gastrointestinal Function and Ulcerationa. Ann N Y Acad Sci 553, 269–285. https://doi.org/10.1111/J.1749-6632.1989.TB54495.X

Takada, K., Zhu, D., Bird, G.H., Sukhdeo, K., Zhao, J.J., Mani, M., Lemieux, M., Carrasco, D.E., Ryan, J., Horst, D., Fulciniti, M., Munshi, N.C., Xu, W., Kung, A.L., Shivdasani, R.A., Walensky, L.D., Carrasco, D.R., 2012. Targeted disruption of the BCL9/β-catenin complex inhibits oncogenic Wnt signaling. Sci Transl Med 4, 148ra117-148ra117. https://doi.org/10.1126/scitranslmed.3003808

Tang, X., Woodward, T., Amar, S., 2009. A PTP4A3 peptide PIMAP39 modulates TNF-alpha levels and endotoxic shock. J Innate Immun 2, 43–55. https://doi.org/10.1159/000235685

Tao, G., Chityala, P.K., 2021. Epidermal growth factor receptor inhibitor-induced diarrhea: clinical incidence, toxicological mechanism, and management. Toxicol Res (Camb) 10, 476–486. https://doi.org/10.1093/TOXRES/TFAB026

Tebbutt, N.C., Wilson, K., Gebski, V.J., Cummins, M.M., Zannino, D., Van Hazel, G.A., Robinson, B., Broad, A., Ganju, V., Ackland, S.P., Forgeson, G., Cunningham, D.,

Saunders, M.P., Stockler, M.R., Chua, Y., Zalcberg, J.R., Simes, R.J., Price, T.J., Price, T., Coates, A., O'Connell, D., Brown, C., Hague, W., France, A., Hicks, S., James, R., Masson, R., O'Connell, R., Pike, R., Shoulder, J., Stevens, L., Tunney, V., Vachan, B., Wong, N., Ackland, S., Moylan, E., Strickland, A., Abdi, E., Ransom, D., Lowenthal, R., Marx, G., Nayagam, S.S., Shannon, J., Goldstein, D., Karapetis, C., Blum, R., Eek, R., Ward, R., Pavlakis, N., Wilcken, N., Burns, I., Wyld, D., Underhill, C., Claringbold, P., Liauw, W., Clingan, P., Jefford, M., Horvath, L., McKendrick, J., Chong, G., Boyce, A., Cassidy, J., Kirsten, F., Clarke, S., Guo, Y., Innes-Rowe, J., Smith, A., Williams, J., Tournier, E., Maliepaard, S., Vitullo, E., Humm, G., Nguyen, V., Midolo, P., Chorlton, C., McDonald, L., Oliver, L., Sjursen, A.M., Inglis, C., Marafioti, T., McCourt, J., Howard, J., Richards, A., Provis, A., Rundle, A., Whatman, A., Emmett, L., Raymond, B., Byrne, S., Withers, E., Campbell, J., Hodgkins, C., Szwajcer, M., Parker, S., Welby, S., Page, F., Corker, M., Wykes, R., Goss, C., Whitney, S., Oates, J., Soulis, E., Hoque, M., Gebbie, C., 2010. Capecitabine, bevacizumab, and mitomycin in first-line treatment of metastatic colorectal cancer: Results of the Australasian Gastrointestinal Trials Group randomized phase III MAX study. J Clin Oncol 28, 3191–3198. https://doi.org/10.1200/JCO.2009.27.7723

The Haplotype Reference consortium: McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C.M., Gillies, C.E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J.C., Boomsma, D., Branham, K., Breen, G., Brummett, C.M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F.S., Corbin, L.J., Smith, G.D., Dedoussis, G., Dorr, M., Farmaki, A.E., Ferrucci, L., Forer, L., Fraser, R.M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O.L., Hveem, K., Kretzler, M., Lee, J.C., McGue, M., Meitinger, T., Melzer, D., Min, J.L., Mohlke, K.L., Vincent, J.B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, J.B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P.E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L.H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M.G., Wilson, J.F., Frayling, T., de Bakker, P.I., Swertz, M.A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C.,

Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C.A., Myers, R.M., Boehnke, M., McCarthy, M.I., Durbin, R., 2016. A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics, 48(10), 1279–1283. https://doi.org/10.1038/ng.3643

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, 205–216. https://doi.org/10.1093/JNCI/92.3.205

Therneau, T., Grambsch, P., 2000. Modeling Survival Data: Extending the Cox Model. https://CRAN.R-project.org/package=survival

Thiagarajah, J.R., Donowitz, M., Verkman, A.S., 2015. Secretory diarrhoea: mechanisms and emerging therapies. Nat Rev Gastroenterol Hepatol 12, 446–457. https://doi.org/10.1038/NRGASTRO.2015.111

Thomas, F., Motsinger-Reif, A.A., Hoskins, J.M., Dvorak, A., Roy, S., Alyasiri, A., Myerson, R.J., Fleshman, J.W., Tan, B.R., McLeod, H.L., 2011. Methylenetetrahydrofolate reductase genetic polymorphisms and toxicity to 5FU-based chemoradiation in rectal cancer. Br J Cancer 105, 1654. https://doi.org/10.1038/BJC.2011.442

Tol, J., Koopman, M., Cats, A., Rodenburg, C.J., Creemers, G.J.M., Schrama, J.G., Erdkamp, F.L.G., Vos, A.H., van Groeningen, C.J., Sinnige, H.A.M., Richel, D.J., Voest, E.E., Dijkstra, J.R., Vink-Börger, M.E., Antonini, N.F., Mol, L., van Krieken, J.H.J.M., Dalesio, O., Punt, C.J.A., 2009. Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. N Engl J Med 360, 563–572. https://doi.org/10.1056/NEJMOA0808268

Toma, M., Beluşică, L., Stavarachi, M., Apostol, P., Spandole, S., Radu, I., Cimponeriu, D., 2012. Rating the environmental and genetic risk factors for colorectal cancer. J Med Life 5, 152. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6880217/

Trikalinos, T.A., Salanti, G., Zintzaras, E., Ioannidis, J.P.A., 2008. Meta-Analysis Methods. Adv Genet 60, 311–334. https://doi.org/10.1016/S0065-2660(07)00413-0

Tsepilov, Y.A., Shin, S.Y., Soranzo, N., Spector, T.D., Prehn, C., Adamski, J., Kastenmüller, G., Wang-Sattler, R., Strauch, K., Gieger, C., Aulchenko, Y.S., Ried,

J.S., 2015. Nonadditive Effects of Genes in Human Metabolomics. Genetics 200, 707–718. https://doi.org/10.1534/GENETICS.115.175760

Turner, S.D., 2018. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J Open Source Softw 3, 731. https://doi.org/10.21105/joss.00731

Urbanek, S., 2013. png: Read and write PNG images. https://CRAN.R-project.org/package=png

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, 1408–1417. https://doi.org/10.1056/NEJMOA0805019/SUPPL_FILE/NEJM_VAN_CUTSEM_1408SA1.PDF

van Cutsem, E., Peeters, M., 1998. Irinotecan monotherapy in the treatment of colorectal cancers: results of phase II trials. Bull Cancer 33–37. https://pubmed.ncbi.nlm.nih.gov/9932082/

van Cutsem, E., Peeters, M., Siena, S., Humblet, Y., Hendlisz, A., Neyns, B., Canon, J.L., van Laethem, J.L., Maurel, J., Richardson, G., Wolf, M., Amado, R.G., 2007. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. J Clin Oncol 25, 1658–1664. https://doi.org/10.1200/JCO.2006.08.1620

van Cutsem, E., Twelves, C., Cassidy, J., Allman, D., Bajetta, E., Boyer, M., Bugat, R., Findlay, M., Frings, S., Jahn, M., McKendrick, J., Osterwalder, B., Perez-Manga, G., Rosso, R., Rougier, P., Schmiegel, W.H., Seit, J.F., Thompson, P., Vieitez, J.M., Weitzel, C., Harper, P., 2001. Oral capecitabine compared with intravenous fluorouracil plus leucovorin in patients with metastatic colorectal cancer: Results of a large phase III study. J Clin Oncol 19, 4097–4106. https://doi.org/10.1200/jco.2001.19.21.4097

van Kuilenburg, A.B.P., Maring, J.G., Schalhorn, A., Terborg, C., Schmalenberg, H., Behnke, D., Schwabe, W., Jabschinsky, K., Hausler, P., 2008. Pharmacokinetics of 5-fluorouracil in patients heterozygous for the IVS14+1G > A mutation in the dihydropyrimidine dehydrogenase gene. Nucleosides Nucleotides Nucleic Acids 27, 692–698. https://doi.org/10.1080/15257770802145009

Varma, A., Mathaiyan, J., Shewade, D., Dubashi, B., Sunitha, K., 2020. Influence of ABCB-1, ERCC-1 and ERCC-2 gene polymorphisms on response to capecitabine and oxaliplatin (CAPOX) treatment in colorectal cancer (CRC) patients of South India. J Clin Pharm Ther 45, 617–627. https://doi.org/10.1111/JCPT.13166

Verlouw, J.A.M., Clemens, E., de Vries, J.H., Zolk, O., Verkerk, A.J.M.H., am Zehnhoff-Dinnesen, A., Medina-Gomez, C., Lanvers-Kaminsky, C., Rivadeneira, F., Langer, T., van Meurs, J.B.J., van den Heuvel-Eibrink, M.M., Uitterlinden, A.G., Broer, L., 2021. A comparison of genotyping arrays. Eur J Hum Genet 2021 29:11 29, 1611–1624. https://doi.org/10.1038/s41431-021-00917-7

Vichaya, E.G., Chiu, G.S., Krukowski, K., Lacourt, T.E., Kavelaars, A., Dantzer, R., Heijnen, C.J., Walker, A.K., 2015. Mechanisms of chemotherapy-induced behavioral toxicities. Front Neurosci 9, 131. https://doi.org/10.3389/FNINS.2015.00131

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 101, 5–22. https://doi.org/10.1016/J.AJHG.2017.06.005

Vitsios, D., Petrovski, S., 2020. Mantis-ml: Disease-Agnostic Gene Prioritization from High-Throughput Genomic Screens by Stochastic Semi-supervised Learning. Am J Hum Genet 106, 659. https://doi.org/10.1016/J.AJHG.2020.03.012

Vreken, P., van Kuilenburg, A.B.P., Meinsma, R., van Gennip, A.H., 1997. Dihydropyrimidine dehydrogenase (DPD) deficiency: identification and expression of missense mutations C29R, R886H and R235W. Hum Genet 101, 333–338. https://doi.org/10.1007/S004390050637

Vuik, F.E.R., Nieuwenburg, S.A.V., Bardou, M., Lansdorp-Vogelaar, I., Dinis-Ribeiro, M., Bento, M.J., Zadnik, V., Pellisé, M., Esteban, L., Kaminski, M.F., Suchanek, S., Ngo, O., Májek, O., Leja, M., Kuipers, E.J., Spaander, M.C.W., 2019. Original article: Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. Gut 68, 1820. https://doi.org/10.1136/GUTJNL-2018-317592

Wang, H., Wang, Z., Rani, P.L., Fu, X., Yu, W., Bao, F., Yu, G., Li, J., Li, L., Sun, L., Yue, Z., Zhao, Q., Pan, Q., Cao, J., Wang, C., Chi, X., Wang, Y., Yang, Q., Mi, Z., Liu, H., Zhang, F., 2017. Identification of PTPN22 , ST6GAL1 and JAZF1 as psoriasis risk genes demonstrates shared pathogenesis between psoriasis and diabetes. Exp Dermatol 26, 1112–1117. https://doi.org/10.1111/exd.13393

Wang, S., Qian, F., Zheng, Y., Ogundiran, T., Ojengbede, O., Zheng, W., Blot, W., Nathanson, K.L., Hennis, A., Nemesure, B., Ambs, S., Olopade, O.I., Huo, D., 2018.

Genetic variants demonstrating flip-flop phenomenon and breast cancer risk prediction among women of African ancestry. Breast Cancer Res Treat 168, 703–712. https://doi.org/10.1007/s10549-017-4638-1

Wang, X., Goldstein, D.B., 2020. Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease. Am J Hum Genet 106, 215–233. https://doi.org/10.1016/j.ajhg.2020.01.012

Wardle, M., Spencer, A., 2017. Implementation of SNOMED CT in an online clinical database. Future Healthc J 4, 126–130. https://doi.org/10.7861/FUTUREHOSP.4-2-126

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, 631–639. https://doi.org/10.1016/S1470-2045(14)70106-8

Weedon, M., Jackson, L., Harrison, J., Ruth, K., Tyrrell, J., Hattersley, A., Wright, C., 2021. Use of SNP chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation. BMJ 372, n214. https://doi.org/10.1136/BMJ.N214

Wei, X., McLeod, H.L., McMurrough, J., Gonzalez, F.J., Fernandez-Salguero, P., 1996. Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. J Clin Investig 98, 610–615. https://doi.org/10.1172/JCI118830

Wickham, H., 2011. The Split-Apply-Combine Strategy for Data Analysis. J Stat Softw, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org

Wickham, H., Hester, J., 2020. readr: Read Rectangular Text Data. https://CRAN.R-project.org/package=readr

Wickham, H., Hester, J., Chang, W., Bryan, J., 2021a. devtools: Tools to Make Developing R Packages Easier. https://CRAN.R-project.org/package=devtools

Wickham, H., François, R., Henry, L., Müller, K., 2021b. dplyr: A Grammar of Data Manipulation. https://CRAN.R-project.org/package=dplyr

Wickham, H., 2021c. tidyr: Tidy Messy Data. https://CRAN.R-project.org/package=tidyr

Wolin, K.Y., Yan, Y., Colditz, G.A., Lee, I.M., 2009. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer 100, 611–616. https://doi.org/10.1038/SJ.BJC.6604917

Won, H.H., Lee, J., Park, J.O., Park, Y.S., Lim, H.Y., Kang, W.K., Kim, J.W., Lee, S.Y., Park, S.H., 2012. Polymorphic markers associated with severe oxaliplatin-induced, chronic peripheral neuropathy in colon cancer patients. Cancer 118, 2828–2836. https://doi.org/10.1002/CNCR.26614

Wu, Q., Dolnick, B.J., 2003. Detection of Thymidylate Synthase Modulators by a Novel Screening Assay. Mol Pharmacol 63, 167–173. https://doi.org/10.1124/MOL.63.1.167

Xie, Y.H., Chen, Y.X., Fang, J.Y., 2020. Comprehensive review of targeted therapy for colorectal cancer. Signal Transduction and Targeted Therapy 2020 5:1 5, 1–30. https://doi.org/10.1038/s41392-020-0116-z

Xu, J., Song, P., Nakamura, S., Miller, M., Barone, S., Alper, S.L., Riederer, B., Bonhagen, J., Arend, L.J., Amlal, H., Seidler, U., Soleimani, M., 2009. Deletion of the chloride transporter Slc26a7 causes distal renal tubular acidosis and impairs gastric acid secretion. J Biol Chem 284, 29470–29479. https://doi.org/10.1074/jbc.M109.044396

Xu, Y., Villalona-Calero, M.A., 2002. Irinotecan: mechanisms of tumor resistance and novel strategies for modulating its activity. Annals of Oncology 13, 1841–1851. https://doi.org/10.1093/ANNONC/MDF337

Yamaguchi, K., Ishigaki, K., Suzuki, A., Tsuchida, Y., Tsuchiya, H., Sumitomo, S., Nagafuchi, Y., Miya, F., Tsunoda, T., Shoda, H., Fujio, K., Yamamoto, K., Kochi, Y., 2022. Splicing QTL analysis focusing on coding sequences reveals mechanisms for disease susceptibility loci. Nat Commun 13. https://doi.org/10.1038/S41467-022-32358-1

Yamamoto, H., Morise, K., Kusugami, K., Furusawa, A., Konagaya, T., Nishio, Y., Kaneko, H., Uchida, K., Nagai, H., Mitsuma, T., Nagura, H., 1996. Abnormal neuropeptide concentration in rectal mucosa of patients with inflammatory bowel disease. J Gastroenterol 31, 525–532. https://doi.org/10.1007/BF02355052

Yang, T., Lapinski, P.E., Zhao, H., Zhou, Q., Zhang, H., Raghavan, M., Liu, Y., Zheng, P., 2005. A rare transporter associated with antigen processing polymorphism overpresented in HLA low colon cancer reveals the functional significance of the

signature domain in antigen processing. Clinical Cancer Research 11, 3614–3623. https://doi.org/10.1158/1078-0432.CCR-04-1804

Yasukawa, Z., Sato, C., Kitajima, K., 2005. Inflammation-dependent changes in α2,3-, α2,6-, and α2,8-sialic acid glycotopes on serum glycoproteins in mice. Glycobiology 15, 827–837. https://doi.org/10.1093/glycob/cwi068

Yasuno, H., Kurasawa, M., Yanagisawa, M., Sato, Y., Harada, N., Mori, K., 2013. Predictive markers of capecitabine sensitivity identified from the expression profile of pyrimidine nucleoside-metabolizing enzymes. Oncol Rep 29, 451–458. https://doi.org/10.3892/or.2012.2149

Ye, F., Liu, Z., Tan, A., Liao, M., Mo, Z., Yang, X., 2013. XRCC1 and GSTP1 polymorphisms and prognosis of oxaliplatin-based chemotherapy in colorectal cancer: a meta-analysis. Cancer Chemother Pharmacol 71, 733–740. https://doi.org/10.1007/S00280-012-2067-8

Yen, J.L., McLeod, H.L., 2007. Should DPD analysis be required prior to prescribing fluoropyrimidines? Eur J Cancer 43, 1011–1016. https://doi.org/10.1016/j.ejca.2007.01.030

Yokokawa, T., Kawakami, K., Mae, Y., Sugita, K., Watanabe, H., Suzuki, K., Suenaga, M., Mizunuma, N., Yamaguchi, T., Hama, T., 2015. Risk factors exacerbating hand-foot skin reaction induced by capecitabine plus oxaliplatin with or without bevacizumab therapy. Annals of Pharmacotherapy 49, 1120–1124. https://doi.org/10.1177/1060028015594451

Zaanan, A., Dumont, L.M., Loriot, M.A., Taieb, J., Narjoz, C., 2014. A case of 5FU-related severe toxicity associated with the p.Y186C DPYD variant. Clin Pharmacol Ther 95, 136. https://doi.org/10.1038/CLPT.2013.183

Zalcberg, J., Kerr, D., Seymour, L., Palmer, M., 1998. Haematological and non-haematological toxicity after 5-fluorouracil and leucovorin in patients with advanced colorectal cancer is significantly associated with gender, increasing age and cycle number. Tomudex International Study Group. Eur J Cancer 34, 1871–1875. https://doi.org/10.1016/S0959-8049(98)00259-7

Zaykin, D. v., Shibata, K., 2008. Genetic Flip-Flop without an Accompanying Change in Linkage Disequilibrium. Am J Hum Genet 82, 794. https://doi.org/10.1016/J.AJHG.2008.02.001

Zeggini, E., Ioannidis, J.P.A., 2009. Meta-analysis in genome-wide association studies. Pharmacogenomics 10, 191. https://doi.org/10.2217/14622416.10.2.191

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., la Manno, G., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K.D., Hjerling-Leffler, J., Arenas, E., Ernfors, P., Marklund, U., Linnarsson, S., 2018. Molecular Architecture of the Mouse Nervous System. Cell 174, 999. https://doi.org/10.1016/J.CELL.2018.06.021

Zhang, M., Qi, T., Yang, L., Kolarich, D., Heisterkamp, N., 2022. Multi-Faceted Effects of ST6Gal1 Expression on Precursor B-Lineage Acute Lymphoblastic Leukemia. Front Oncol 12, 688. https://doi.org/10.3389/FONC.2022.828041/BIBTEX

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, 116487. https://doi.org/10.1371/JOURNAL.PONE.0116487

Zhou, X., Kinlough, C.L., Hughey, R.P., Jin, M., Inoue, H., Etling, E., Modena, B.D., Kaminski, N., Bleecker, E.R., Meyers, D.A., Jarjour, N.N., Trudeau, J.B., Holguin, F., Ray, A., Wenzel, S.E., 2019. Sialylation of MUC4β N-glycans by ST6GAL1 orchestrates human airway epithelial cell differentiation associated with type-2 inflammation. JCI Insight 4. https://doi.org/10.1172/JCI.INSIGHT.122475

Zielinski, C., Lang, I., Beslija, S., Kahan, Z., Inbar, M.J., Stemmer, S.M., Anghel, R., Vrbanec, D., Messinger, D., Brodowicz, T., 2016. Predictive role of hand-foot syndrome in patients receiving first-line capecitabine plus bevacizumab for HER2-negative metastatic breast cancer. Br J Cancer 114, 163–170. https://doi.org/10.1038/BJC.2015.419

Zimmerman, M.W., Homanics, G.E., Lazo, J.S., 2013. Targeted Deletion of the Metastasis-Associated Phosphatase Ptp4a3 (PRL-3) Suppresses Murine Colon Cancer. PLoS One 8. https://doi.org/10.1371/journal.pone.0058300

Zirpoli, G.R., McCann, S.E., Sucheston-Campbell, L.E., Hershman, D.L., Ciupak, G., Davis, W., Unger, J.M., Moore, H.C.F., Stewart, J.A., Isaacs, C., Hobday, T.J., Salim, M., Hortobagyi, G.N., Gralow, J.R., Budd, G.T., Albain, K.S., Ambrosone, C.B., 2017. Supplement Use and Chemotherapy-Induced Peripheral Neuropathy in a Cooperative Group Trial (S0221): The DELCaP Study. J Natl Cancer Inst 109. https://doi.org/10.1093/jnci/djx098

Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., Lander, E.S., 2014. Searching for missing heritability:

Designing rare variant association studies. Proc Natl Acad Sci USA 111, E455–E464. https://doi.org/10.1073/PNAS.1322563111