IDENTIFICATION OF GENETIC FACTORS PREDISPOSING TO CHEMOTHERAPY AND IMMUNOTHERAPY TOXICITY

by

IK SHIN CHIN

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences

College of Medical and Dental Sciences

University of Birmingham

July 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Introduction

Serious Common Terminology Criteria for Adverse Events (CTCAE) grade 3-4 toxicities can cause substantial morbidity and negatively impact patients' quality of life. The ability to predict patients' toxicity risk from systemic anti-cancer therapy can help personalise treatment. Germline genetic variation has been shown to be a key determinant of toxicity risk. My thesis aimed to identify genetic variants predisposing towards chemotherapy and immune checkpoint inhibitor (ICI) toxicity. I performed genome wide association studies (GWAS) and fine mapping studies of both fluoropyrimidine-induced toxicity and ICI-induced immune related adverse events (irAEs). I also analysed candidate variants and performed functional analysis for one locus.

Methods

Per cycle CTCAE graded toxicity data was obtained from 6459 colorectal cancer patients who received fluoropyrimidine chemotherapy from four clinical trials. CTCAE graded irAE were analysed from a discovery cohort of 599 patients on ICIs. 103 cases of serious irAE and 2497 controls from UK Biobank formed a replication cohort. Tagging single nucleotide polymorphism (SNP) arrays were used for genotyping and genome-wide imputation was performed. Association testing and meta-analyses were conducted for common capecitabine and ICI induced toxicity variables. Conditional and joint association analysis was done to identify top independent SNPs that were tested in polygenic risk score (PRS) analyses. Preliminary functional experiments were also conducted for a gene implicated in fluoropyrimidine induced cardiotoxicity.

Results

In relation to toxicities induced by fluoropyrimidines, I identified four SNPs associated with early CTCAE grade 3-4 capecitabine toxicity at GWAS significance ($P<5x10^{-8}$) and a further two novel SNPs associated with cardiovascular toxicity under the additive genetic model. I also generated data that did not validate the current literature that *MIR27A* variant rs895819 is only associated with toxicity in *DPYD* no function/low function allele carriers.

In relation to ICI-induced irAEs, I identified eight independent SNPs associated with any serious irAE at suggestive significance ($P<5x10^{-6}$) under the additive model. Seven and 12 SNPs were associated with ≥grade 3 GI and hepatitis irAEs at $P<5x10^{-6}$. A diabetes PRS was significantly associated with hyperglycaemic irAEs events. Four SNPs associated with ≥grade 3 endocrine, GI and hepatitis irAEs were replicated in the validation phase at nominal significance ($P<0.05$), although none remained genome-wide significant following meta-analysis. SNPs associated with any serious irAE, immune-related hepatitis and a novel toxicity score is mapped to the *ZNF701* gene at suggestive significance.

Conclusion

We identified potential genetic variants that can contribute towards predictive risk models of toxicity. Use of these markers in practice should be balanced against markers of efficacy to maximise treatment benefit.

Acknowledgements

Firstly I would like to thank Dr Claire Palles who is my lead supervisor for my PhD. She has helped me tremendously throughout my PhD, particularly coming from a clinical background, I am grateful for all her patience and hard work in helping me develop various scientific and research skills and knowledge that would be help me in future academic projects. She has always looked after my training needs and career progression, and I am very lucky to be part of her group. I would also like to thank my co-supervisors Professor Gary Middleton for ensuring my learning needs are met and for being a mentor as a clinical academic and both him and Dr Neil Steven have also helped to make improvements to the protocol and study questionnaire and recruitment to the ICI Genetics Study.

I would like to thank my lab group members Sara Galavotti, Helen Curley and Laura Chegwidden for their continuous help and support. I am grateful for Sara who has taught me the wet lab techniques including gel electrophoresis, DNA and RNA extraction, tissue culture work, gene expression analysis and many more. Helen and Laura have helped me develop skills in script writing and bioinformatic analysis that has enabled me to carry out the analyses for my project. I am privileged to be in a very supportive environment within the group.

I would like to thank my family and friends for always supporting me, especially during the last few challenging years during the pandemic. I am grateful for the continuous encouragement from my parents who have always had my back. I have been very lucky to have a very understanding and supportive new husband (thankfully managed to have our wedding despite all the COVID uncertainties) who have been there for me through thick and thin and who has patiently helped with the housework and cooking as well as looking after my well-being throughout these years.

I would also like to thank all the patients who have participated in the ICI GENETICS study and the study's consortium research team members who have continuously helped support recruitment to our study. Thank you to the study sponsor and NIHR team who has enabled the study to run smoothly. Due to their hard work and contribution we were able to surpass the expected recruitment numbers despite being held back by the COVID-19 pandemic. Thank you to our collaborators for their contribution, support and feedback in our work to assure good quality research outputs are achieved.

Finally, I would like to express my heartfelt gratitude towards Cancer Research UK Birmingham Centre for awarding me this Clinical PhD fellowship grant and allowing me to embark on this academic journey. I have gained valuable opportunities and experiences and met inspiring individuals that have motivated me to continue to pursue a clinical academic career.

COVID-19 DECLARATION

The ICI GENETICS study which is the main study where majority of my data analysis for my thesis would be based on was significantly impacted during the COVID-19 pandemic. During the peak of the first wave in 2020, the study had to be paused temporarily which affected all study recruitment and activity. This also halted the opening of new study sites and hence recruitment was slow to recover when the study was allowed to reopen. As a result, this created a setback to the anticipated sample size we were planning to achieve in order to perform an adequately powered genome-wide association analysis as part of the main aims for my thesis.

To overcome this and increase the discovery cohort size, we collaborated with other study groups from Kings, Liverpool, Leeds, Newcastle, and Oxford which also collected immunotherapy toxicity data and whole blood DNA for germline analyses. We also had to amend our laboratory risk assessment protocol on the processing of our study samples received. This meant that all study samples had to heat incubated in a biosafety level 2 cabinet to inactivate any potential COVID-19 viral particles and protect the user from the risk of infection. This has subsequently lengthened the time required to process and store the samples with additional resources required. An extension to my PhD programme has also been approved by the college and Cancer Research UK Birmingham Centre as the funder for my fellowship.

During the pandemic when the laboratories and university was closed, I conducted a systematic review and meta-analysis on the influence of clinical characteristics and anti-cancer therapy on outcomes from COVID-19 infection in cancer patients. This has helped me develop and learn new statistical analytical skills which I can also apply in future research projects.

Together with other members of my lab group as co-authors, we published this work as a pre-print in medRxiv[1].

Despite the challenges faced as a consequence of the pandemic, I believe the work presented in this thesis has still make a significant contribution in this field. The data that has been collected and analysed will help to guide further large research studies in a pooled or meta-analysis and guide further functional work to test the implications of potential variants predictive of toxicity.

Table of Contents

Abstract *2*

Acknowledgements *4*

COVID-19 DECLARATION *6*

List of Figures *17*

List of Tables *21*

List of Abbreviations *26*

CHAPTER 1: INTRODUCTION *1*

1.1 Cancer treatment **2**

1.2 Fluoropyrimidine chemotherapy **2**

1.2.1 Metabolism of 5-FU chemotherapy 2

1.2.2 Activation pathway of capecitabine 4

1.3 The immune checkpoint inhibitors **5**

1.4 Toxicities related to cancer treatment **7**

1.4.1 Fluoropyrimidine-induced toxicity 8

1.4.2 Cardiovascular toxicity 8

1.4.2.1 Pre-clinical studies assessing effects of 5-FU on cardiac cells 9

1.5 ICI induced toxicity **10**

1.5.1 The spectrum and challenges of immune-related adverse events 10

1.5.2 Toxicity profile of different ICI therapies 10

1.5.3 Patient and tumour characteristics that may impact on risk of irAEs 14

1.5.4 Time to onset of irAE 14

1.6 Role of germline genetic variation in determining toxicity risk **15**

1.6.1 Germline determinants of host immunity 17

1.6.1.1 The cancer immunity cycle 17

1.6.1.2 Interplay between host genetics and immuno-modulatory cells and molecules 18
1.6.2 Genetics of Autoimmune Disease 20
1.6.2.1 Association between the MHC locus and autoimmune disease 22
1.6.2.2 Non-MHC susceptibility loci 23
1.6.2.3 Challenges in genetic studies of autoimmune disease will also apply to studies of irAE susceptibility 23
1.6.2.4 ICI in patients with autoimmune disease 24
1.7 Identification of genetic factors predisposing to toxicity 25
1.7.1 Methodology for identifying genetic variants predictive of toxicity 25
1.7.1.1 Genome-wide association studies 25
1.7.1.2 Polygenic risk score analysis 27
1.7.2 Genetic factors predisposing to fluoropyrimidine toxicities 28
1.7.2.1 DPYD gene polymorphisms in 5-FU toxicity screening 28
1.7.2.2 Genetic factors predisposing to capecitabine toxicity 30
1.7.2.3 Association of MIR27A polymorphisms with fluoropyrimidine toxicity 35
1.7.3 Genetic determinants of chemotherapy-induced cardiotoxicity 36
1.7.4 Genetic factors influencing irAE development 37
1.7.4.1 Current evidence of genetic factors influencing irAE development 37
1.8 Association between treatment toxicity and response 44
1.8.1 Genetic factors associated with treatment efficacy 45
1.8.2 Are genetic variants associated with response also predictive of toxicity risk? 47
1.8.3 Developing a predictive model of toxicity for use in clinical practice 47
CHAPTER 2: MATERIALS AND METHODS 52
2.1 Patient samples 53
2.1.1 QUASAR 2 53
2.1.2 COIN 54
2.1.3 SCOT 55
2.1.4 Dutch cohort 55

2.1.5 ICI GENETICS 55
2.1.5.1 Participating Sites 57
2.1.5.2 Sample Collection 58
2.1.5.3 Data Collection 58
2.1.5.4 Sample storage and DNA extraction 59
2.1.6 Discovery phase samples from IO-BASE consortium members 59
2.1.7 Validation cohort 60
2.2 Coding adverse events for analysis 61
2.2.1 Coding capecitabine/5-FU-induced toxicity 61
2.2.2 Coding ICI-induced immune-related adverse events 62
2.3 Statistical analysis 63
2.3.1 File formatting 63
2.3.2 Frequentist based genome-wide association study analysis 63
2.3.3 Analysis of genome-wide association study data using PLINK 64
2.3.4 Analysis of genome-wide association study using the cox proportional hazard ratio 65
2.3.5 Meta-analyses 65
2.3.6 Variant annotation 65
2.3.7 Conditional analyses and joint association tests 66
2.3.8 Gene Set and Tissue Expression Analysis 66
2.3.9 Interrogating publicly available datasets with expression quantitative trait loci (eQTL) and chromatin interaction information 67
2.3.10 Phenome-wide association study (PheWAS) 68
2.3.11 Polygenic risk scores 69
2.3.12 Receiver operating curve analysis 70
2.3.13 Transcriptomic wide association analysis (TWAS) 70
2.3.14 Power calculations 71
2.4 KASP Genotyping 71
2.5 Polymerase chain reaction, PCR clean up, agarose gels and Sanger sequencing 73

2.6 Cell culture experiments assessing the impact of 5-FU and *DGLUCY* knockdown on mouse cardiomyocytes 74

2.6.1 Preparing HL-1 cell media 74

2.6.2 Cell culture 75

2.6.3 Cell counting and plating for metabolite extraction 75

2.6.4 Metabolite harvest and extraction for gas chromatography-mass spectrometry (GC-MS) 76

2.6.5 Treatment of cells with 5-fluorouracil 77

2.6.6 Measuring cell viability 77

2.6.7 Knockdown of *DGLUCY* in cells 78

2.6.7.1 Transfection reactions 78

2.6.7.2 Optimising transfection experiments 78

2.6.7.3 Assessing cell viability of DGLUCY knockdown cells post 5-FU treatment 79

2.6.7.4 Analysing gene expression of DGLUCY knockdown transfected cells post 5-FU treatment 80

2.7 Mitochondrial isolation for detection of D-glutamate 81

2.7.1 Digitonin concentration optimisation 81

2.7.2 Digitonin cell fractionation to isolate mitochondrial from transfected and non-transfected HL-1 cells 82

2.7.2.1 Non-transfected cells 82

2.7.2.2 Transfected cells 83

2.8 RNA extraction 84

2.8.1 Removal of DNA 84

2.8.2 RNA quantification 84

2.9 Reverse transcription 85

2.10 Real-time PCR (RT-qPCR) 85

2.11 DNA extraction 86

2.12 DNA quantification 87

2.13 Whole Genome tagging SNP arrays 88

2.14 Quality controls steps applied to genotyping data 88

2.15 Phasing and Imputation 89

2.16 Analysis of whole genome sequencing data 90

CHAPTER 3: IDENTIFICATION OF GENETIC FACTORS PREDISPOSING TO FLUOROPYRIMIDINE-INDUCED TOXICITY 94

3.1 Introduction 95

3.1.1 Genetic determinants of capecitabine-induced toxicity 96

3.1.2 The role of mir27A in predicting risk of early-onset fluorouracil toxicity 97

3.1.3 Genetic determinants of maximum grade all cycle capecitabine-induced toxicity in the QUASAR2 trial 99

3.1.4 Chapter aims 100

3.2 Results 100

3.2.1 Patient cohort 100

3.2.2 Analysis of previously reported variants for association with early cycle capecitabine induced toxicity in QUASAR2 101

3.2.3 Genome-wide association analysis of capecitabine toxicity in QUASAR2 103

3.2.4 Top SNPs associated with early-onset toxicity using different methods to handle genotype uncertainty 104

3.2.5 Genome-wide association analysis of early-onset capecitabine toxicity 104

3.2.5.1 Comparison of the method score and method expected approach 108

3.2.6 Genome-wide association analysis of capecitabine toxicity across all cycles 110

3.2.7 Meta-analysis GWAS of early ⩾grade 3 capecitabine toxicity 112

3.2.8 Association between *MIR27A* SNPs with early-onset fluoropyrimidine toxicity 121

3.2.8.1 KASP genotyping for both MIR27A SNPs in QUASAR 2 121

3.2.8.2 Association between common MIR27A SNP rs895819 and early and all cycle ⩾grade 3 toxicity in QUASAR2 123

3.2.8.3 Association between rare MIR27A SNP rs11671784 with early cycle and all cycle ⩾grade 3 toxicity in QUASAR2 124

3.2.8.4 Meta-analysis investigating the association between MIR27A SNPS and early ⩾grade 3 toxicity in QUASAR2 and Dutch cohorts 125

3.2.8.5 Meta-analysis of the association between MIR27A SNP rs11671784 and early ⩾grade 3 toxicity in QUASAR2 and Dutch study cohorts 127

3.2.8.6 Stratified analyses between MIR27A rs895819 and early ⩾grade 3 toxicity according to DPYD carrier status 128

3.2.8.7 Sensitivity analyses of MIR27A rs895819 in predicting early toxicity 129

3.2.9 Polygenic risk profiling 130

3.3 Discussion 134

3.3.1 Previously reported variants of ⩾grade 3 toxicity 134

3.3.2 Novel variants associated with ⩾grade 3 toxicity at genome-wide significance 134

3.3.3 Determining toxicity risk using polygenic risk scores and toxicity prediction models 138

3.3.4 Limitations of analyses 139

3.3.5 Implication of findings 140

CHAPTER 4: FUNCTIONAL ANALYSIS OF A GENE ASSOCIATED WITH 5-FU INDUCED CARDIOVASCULAR TOXICITY AND IDENTIFICATION OF GENETIC DETERMINANTS OF CARDIAC TOXICITY 142

4.1 Introduction 143

4.1.1 Genetic markers of fluoropyrimidine-induced cardiotoxicity 143

4.1.2 Pre-clinical studies determining the role of DGLUCY 146

4.1.3 Effects of 5-FU on amino acid metabolism 147

4.1.4 Investigating the role of *DGLUCY* on cardiac toxicity in the presence of 5-FU 148

4.1.5 Chapter aims 149

4.2 Results 149

4.2.1 Functional analysis of *DGLUCY* gene 149

4.2.1.1 Effects of 5-FU chemotherapy on viability of HL-1 cells 149

4.2.1.2 Gene expression of DGLUCY following treatment with 5-FU 152

4.2.1.3 Detection of L-glutamate and D-glutamate by gas chromatography-mass spectrometry (GC-MS) 153
4.2.1.4 D-glutamate is undetectable after treatment of HL-1 cells with 5-FU 157
4.2.1.5 Knockdown of DGLUCY by siRNA transfection in HL-1 cells 157
4.2.1.6 D-glutamate was still undetected after siRNA knockdown of DGLUCY in either mitochondrial or cytoplasmic cell fractions 158
4.2.2 Genetic determinants of fluoropyrimidine cardiotoxicity 162
4.2.2.1 Are variants recommended for testing before treatment with 5-FU associated with cardiovascular toxicity? 162
4.2.2.2 Meta-analysis of summary statistics from three colorectal cancer clinical trials investigating genetic determinants of cardiovascular toxicity following treatment with 5FU-containing regimens 163
4.3 Discussion 168
4.3.1 Determining the functional role of *DGLUCY* in the presence of 5-FU 168
4.3.2 Identifying genetic variants predisposing towards fluoropyrimidine cardiotoxicity 169
4.3.3 Future directions 171
CHAPTER 5: IDENTIFICATION OF GENETIC FACTORS PREDISPOSING TO IMMUNE CHECKPOINT INHIBITOR TOXICITY 172
5.1 Introduction 173
5.1.1 Chapter aims 174
5.2 Results 175
5.2.1 Recruitment to ICI GENETICS 175
5.2.2 Genome-wide association studies of ICI toxicity 177
5.2.2.1 Sample selection and initial genotype quality control for GWAS Discovery Phase 177
5.2.2.2 Toxicity profile of GWAS cohort 180
5.2.2.3 Power calculation 182
5.2.2.4 Association analysis of serious irAEs based on additive genetic models 182
5.2.2.5 Association analysis of individual organ irAEs based on additive genetic models 188
5.2.2.6 Association analysis based on Cox proportional hazard models 208

5.2.3 Association analysis of HLA alleles 211
5.2.4 Replication of previously reported SNPs associated with irAE 213
5.2.5 Polygenic risk score analysis 214
5.2.5.1 PRS analysis based on SNPs associated with autoimmune disease 214
5.2.5.2 PRS analysis based on SNPs associated with immuno-modulatory traits 217
5.2.5.3 PRS analysis based on top SNPs associated with GWAS analysis 217
5.2.6 Validation of top SNPs from GWAS 219
5.2.6.1 In silico evaluation in an external GWAS cohort 219
5.2.6.2 Validation in an independent cohort from ICI GENETICS 223
5.2.7 Transcriptome-wide association analysis (TWAS) for ⩾grade 3 irAE 225
5.2.8 Whole genome sequencing analysis of samples with multiple ⩾grade 3 toxicities 226
5.2.9 Correlation between toxicity and overall survival 231
5.3 Discussion 232
5.3.1 Toxicity profile of our discovery cohort 232
5.3.2 Markers of serious irAEs of any type 232
5.3.3 Markers of individual toxicities 235
5.3.4 Insights from gene-based tests 237
5.3.5 The use of polygenic risk scores to identify patients at risk of irAEs 238
5.3.6 Whole genome sequencing 239
5.3.7 Alternative toxicity grading systems 240
5.3.8 Relationship between ICI efficacy and toxicity 240
5.3.9 Strengths and limitations 242
5.3.10 Future directions 243
CHAPTER 6: DISCUSSION 245
6.1 Toxicity induced by chemotherapy or immunotherapy 246
6.2 Role of germline genetic variation in predicting toxicity 246
6.3 Using GWAS to identify candidate genetic markers of toxicity 248
6.4 Challenges in conducting studies of irAE 250

6.5 Are genetic variants predictive of toxicity associated with better treatment efficacy? 251

6.6 Future direction 253

6.7 Concluding remarks 257

CHAPTER 7: REFERENCES 258

CHAPTER 8: APPENDICES 298

8.1 Appendix 1- Review of germline genetics and ICI toxicity 299

8.2 Appendix 2- CTCAE grading 300

8.3 Appendix 3- Immune-related genes 301

8.4 Appendix 4- Top genetic markers incorporated into the PRS of autoimmune diseases to predict irAE development 304

List of Figures

Figure 1.1 Metabolism pathway of fluoropyrimidine chemotherapy. 4

Figure 1.2 Activation and metabolism pathway of 5-FU oral prodrug capecitabine..... 5

Figure 1.3 Diagram illustrating the activity of the immune checkpoint inhibitors..... 6

Figure 1.4 Cancer immunity cycle 18

Figure 2.1 Flow chart demonstrating the recruitment process for ICI GENETICS 57

Figure 2.2 96-well plates containing siRNA transfected HL-1 cells 79

Figure 2.3 Summary of the pipeline used to process WGS data (work courtesy of Dr Claire Palles) 93

Figure 3.1 Manhattan plots showing results from the GWAS meta-analysis of ⩾grade 3 capecitabine toxicity in QUASAR2 and COIN within the first two treatment cycles 115

Figure 3.2 Genotypic profile of two *MIR27A* SNPs in the QUASAR2 trial cohort according to imputation or KASP genotyping derived genotypes 122

Figure 3.3 AUC-ROC of the top independent SNPs associated with mucositis and haematological toxicity 131

Figure 3.4 AUC-ROC of the top independent SNPS associated with diarrhoea and vomiting toxicity..... 132

Figure 3.5 AUC-ROC of the top independent SNPS associated with hand-foot-syndrome and global toxicity 133

Figure 3.6 Possible predictive model of ⩾grade 3 toxicity risk in patients starting fluoropyrimidine-based therapies..... 139

Figure 4.1 Genetic variants associated with cardiovascular toxicity 144
Figure 4.2 Line plots of HL-1 cells treated with a single dose of 5-FU chemotherapy following MTS assay administration. 150
Figure 4.3 Line plots of HL-1 cells treated with two doses of 5-FU chemotherapy following MTS assay administration. 151
Figure 4.4 Fold difference in gene expression between untreated and treated HL-1 cells. . 153
Figure 4.5 Chiral glutamate on Agilent Chirasil-Val 154
Figure 4.6 GC-MS chromatogram showing the detection of L-pyroglutamate from mouse heart tissue 155
Figure 4.7 GC-MS chromatogram showing the detection of the L-glutamate extracted from HL-1 mouse cardiomyocytes. 156
Figure 4.8 Comparison between L-glutamate and D-glutamate levels between untreated and 5-FU treated HL-1 cells 157
Figure 4.9 Fold difference in gene expression between the SCR cells and cells transfected with 30pmol siRNA targeting *DGLUCY* at 24 and 48 hours post-transfection. 158
Figure 4.10 Images of HL-1 cells tested with different concentrations of digitonin and stained with trypan blue as visualised under the microscope at 40X magnification. 160
Figure 4.11 a) Chromatogram showing the peak areas of D- and L- glutamate isomers extracted from a pool of mitochondrial replicates that were non-transfected. b) Comparison of L-glutamate levels between whole cells and mitochondrial extracts 161
Figure 4.12 Manhattan plot showing results from the GWAS meta-analysis of cardiac toxicity from 3 trial cohorts (QUASAR2, COIN and SCOT) 164
Figure 4.13 Chromatin interaction mapping for chromosome 7 based on GWAS meta-analysis of cardiac toxicity (additive model) 165

Figure 4.14 Manhattan plot showing results from the GWAS meta-analysis of cardiac toxicity from three trial cohorts (QUASAR2, COIN and SCOT) 166

Figure 5.1 Participants recruited to the ICI GENETICS Study (IRAS 237779) up till October 2022 175

Figure 5.2 Consort flow diagram of outputs from the QC steps applied to the genotyping data of samples included in the GWAS analysis of irAE. 178

Figure 5.3 Number of patients with CTCAE graded irAE included in the GWAS analysis. 181

Figure 5.4 Manhattan plot from the genome-wide association analysis identifying variants predictive of serious irAE of any type. 183

Figure 5.5 MAGMA tissue expression analysis of GWAS of serious irAE 186

Figure 5.6 Chromatin interaction mapping for chromosome 19 based on GWAS of serious irAE 188

Figure 5.7 Manhattan plot from the genome-wide association analysis identifying variants predictive of ⩾grade 3 GI irAE 189

Figure 5.8 MAGMA tissue expression analysis from GWAS of ⩾grade 3 GI toxicity 192

Figure 5.9 Manhattan plot from the genome-wide association analysis identifying variants predictive of ⩾grade 3 immune-related hepatitis 194

Figure 5.10 MAGMA tissue expression analysis of GWAS of ⩾grade 3 immune-related hepatitis 197

Figure 5.11 PRS prediction for irAE 216

Figure 5.12 AUC-ROC curve and density plot of the top 8 SNPs associated with serious irAE in the prediction of irAE. 218

Figure 5.13 AUC-ROC curve and density plot of the top 8 SNPs associated with serious irAE in the prediction of irAE in the validation cohort..... 219
Figure 5.14 Unadjusted survival curves for patients with and without serious irAE (n=448) 231
Figure 5.15 Proposed mechanisms of how ZNF701 expression influences irAE development 234

List of Tables

Table 1.1 Types of ICI therapies and their therapeutic indications 7

Table 1.2 Frequency of different irAE according to the ICI treatment type 12

Table 1.3 Genetic variants associated with immuno-modulatory traits in the cancer immunity cycle 19

Table 1.4 Genetic variants associated with common autoimmune disease traits 21

Table 1.5 Four *DPYD* variants included in the screening test and listed in the CPIC® guidelines 29

Table 1.6 Previously reported genetic variants associated with capecitabine-induced toxicities 33

Table 1.7 Genetic variants associated with irAE development from current studies 38

Table 2.1 Eligibility criteria for enrolment to the ICI GENETICS study 56

Table 2.2 Studies of the collaboration that contributed to the initial GWAS of irAE 60

Table 2.3 ICI-induced events reported for each organ-specific irAE variable 62

Table 2.4 KASP genotyping mix components 72

Table 2.5 PCR master mix (10µl reactions) 73

Table 2.6 Timeline in the steps taken to assess cell viability and gene expression of transfected HL-1 cells treated with 5-FU 81

Table 2.7 Contents of assay tubes for DNA quantification 87

Table 3.1 Number of patients with capecitabine-induced toxicities in QUASAR2 101

Table 3.2 Association between previously reported variants with early onset global toxicity during the first two treatment cycles in QUASAR2 102

Table 3.3 Top independent SNPs from GWAS of early-onset ⩾grade 3 capecitabine toxicity using the “score” approach to handle genotype uncertainty under the additive genetic model 105
Table 3.4 Top independent SNPs from GWAS of early ⩾grade 3 capecitabine toxicity using the “expected” approach to handle genotype uncertainty under the additive genetic model 107
Table 3.5 Summary statistics of the top SNPs from the score method approach in both the score and expected method analysis 109
Table 3.6 Top independent SNPs from GWAS of ⩾grade 3 capecitabine toxicity across all treatment cycles using the “expected” approach to handle genotype uncertainty under the additive genetic model 111
Table 3.7 Number of early-onset toxicity cases in both QUASAR2 and COIN trial cohorts... 112
Table 3.8 Top SNPs identified from the meta-analysis of QUASAR2 and COIN datasets of early ⩾grade 3 capecitabine toxicity under the additive genetic model (total N=1928)..... 117
Table 3.9 Expression quantitative trait loci (eQTL) data from top SNPs from the GWAS meta-analysis of early capecitabine toxicity 119
Table 3.10 Top 10 gene sets from MAGMA gene-set analysis from GWAS meta-analysis of early ⩾grade 3 global capecitabine toxicity 120
Table 3.12 Summary statistics of the association between rs895819 T>C variant and ⩾grade 3 capecitabine-induced toxicities using both imputed and KASP genotyping data 123
Table 3.13 Summary statistics of the association between rs11671784 G>A variant and any (global) ⩾grade 3 capecitabine toxicity using KASP genotyping data 124
Table 3.14 Number of cases in the Dutch cohort with ⩾grade 3 toxicities in treatment cycles 1 and 2 126

Table 3.15 Meta-analysis of rs895819 with early ⩾grade 3 toxicity from QUASAR2 and the Dutch study cohort (additive model) 126

Table 3.16 Meta-analysis of rs895819 with early ⩾grade 3 toxicity from QUASAR2 and the Dutch study cohort (recessive model) 127

Table 3.17 Meta-analysis of rs11671784 with early ⩾grade 3 toxicity from QUASAR2 and the Dutch study cohort 128

Table 3.18 Meta-analysis of rs895819 with early ≥grade 3 diarrhoea, GI and global toxicity in QUASAR2 and Dutch cohorts according to *DPYD* carrier status 129

Table 3.19 Predictive ability of MIR27A rs895819 and CPIC DPYD risk variants in determining the risk of early ⩾grade 3 capecitabine toxicity..... 129

Table 3.20 Performance of rs2477888 in the meta-analysis of early capecitabine toxicity.. 136

Table 4.1 Association of *DPYD* risk variants with cardiac toxicity under the additive model 162

Table 4.2 Three large colorectal cancer trials included in the meta-analysis of cardiotoxicity 163

Table 4.3 Top SNPs identified from GWAS meta-analysis of fluoropyrimidine induced cardiac toxicity from 3 trial cohorts under the additive genetic model 164

Table 4.4 Top SNPs identified from GWAS meta-analysis of fluoropyrimidine induced cardiac toxicity from three trial cohorts under the recessive genetic model..... 167

Table 5.1 Baseline characteristics of patients recruited to the ICI GENETICS study 176

Table 5.2 Baseline characteristics of patients included in the GWAS of irAE who received ICI treatment 179

Table 5.3 Top independent SNPs identified from GWAS of serious irAE under the frequentist additive genetic model..... 184

Table 5.4 Top 10 gene sets identified from the GWAS of any serious irAE 185

Table 5.5 Expression quantitative trait loci (eQTL) data from top SNPs associated with the presence of any serious irAE 187

Table 5.6 Top independent SNPs identified from GWAS of ⩾grade 3 GI irAE under the frequentist additive genetic model 190

Table 5.7 Top 10 gene sets from MAGMA gene-set analysis for GWAS of ⩾grade 3 GI toxicity 191

Table 5.8 Expression quantitative trait loci (eQTL) data of the top SNPs associated with the presence of ⩾grade 3 GI irAE 193

Table 5.9 Top independent SNPs identified from GWAS of ⩾grade 3 hepatitis irAE under the frequentist additive genetic model 195

Table 5.10 Top 10 gene sets from MAGMA gene-set analysis for GWAS of ⩾grade 3 hepatitis 196

Table 5.11 Expression quantitative trait loci (eQTL) data of top SNPs associated with the presence of ⩾grade 3 immune-related hepatitis 198

Table 5.12 Top independent SNPs identified from GWAS of ICI-induced endocrine, pneumonitis, skin and musculoskeletal toxicities under the frequentist additive genetic model 201

Table 5.13 Expression quantitative trait loci (eQTL) data of top SNPs associated with the presence of ⩾grade 3 endocrine and ⩾grade 2 pneumonitis, skin and MSK irAE 204

Table 5.14 Top SNPs associated with cumulative OST scores under the additive model that reached GWAS significance (P x $5x10^{-8}$) 207

Table 5.15 Performance of the top SNPs associated with serious irAE and OST score 208

Table 5.16 Top independent SNPs identified from GWAS of serious irAE under the Cox proportional hazard model 210
Table 5.17 Top genetic variants within the chromosome 6 MHC region that are associated with the development of serious irAE 212
Table 5.18 Summary statistics of previously reported SNPs in their association with any serious irAE in our discovery study cohort 213
Table 5.19 Summary statistics from PRS analysis of top SNPS correlated with different AD traits in predicting any serious irAE and associated ⩾grade 3 end-organ toxicities. 215
Table 5.20 Summary statistics of the top SNPs identified from GWAS of irAE in our discovery cohort and a US validation cohort 221
Table 5.21 Summary statistics of the top markers associated with any serious irAE from the discovery GWAS in the discovery and validation cohorts and meta-analysis of both cohorts 223
Table 5.22 Summary statistics of the top markers associated with ⩾grade 3 GI irAE from the discovery GWAS in the discovery and validation cohorts and meta-analysis of both cohorts 224
Table 5.23 Summary statistics of the top markers associated with ⩾grade 3 immune-related hepatitis GI irAE from the discovery GWAS in the discovery and validation cohorts and meta-analysis of both cohorts 225
Table 5.24 ⩾Grade 3 irAE experienced by the 8 patients where WGS was performed on their samples 226
Table 5.25 Types of VEP filter to the annotated variants 227
Table 5.26 Top rare variants (allele frequencies <1%) from patients with multiple ⩾grade 3 irAE 229

List of Abbreviations

AD	Autoimmune Disease
AE	Adverse Event
ALT	Alternate
AS	Ankylosing Spondylitis
AUC	Area Under the Curve
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
CI	Confidence interval
COJO	Conditional and Joint analysis
CNIO	Spanish National Cancer Research Centre
CPIC	Clinical Pharmacogenomics Implementation Consortium
CRC	Colorectal Cancer
CRN	Clinical Research Network
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
CTLA4	Cytotoxic T lymphocyte antigen
DPYD	Dihydropyrimidine Dehydrogenase
EBI	European Bioinformatic Institute
EDTA	Ethylenediaminetetraacetic Acid
eQTL	Expression Quantitative Trait Loci
5-FU	5-Fluorouracil
FDR	False Discovery Rate
GC-MS	Gas Chromatography-Mass Spectrometry
GI	Gastrointestinal

GO	Gene Ontology
GSA	Global Screening Array
GTEx	Genome-Tissue Expression
GWAS	Genome Wide Association Study
HFS	Hand-foot-syndrome
HLA	Human Leucocyte Antigen
HR	Hazard Ratio
HRA	Health Research Authority
HRC	Haplotype Reference Consortium
HWE	Hardy Weinberg Equilibrium
IBD	Inflammatory Bowel Disease
ICI	Immune Checkpoint Inhibitor
IRAE	Immune-related Adverse Event
IRAS	Integrated Research Application System
KASP	Kompetitive allele specific PCR
LD	Linkage Disequilibrium
LOF	Loss-Of-Function
LRT	Likelihood Ratio Test
MAF	Minor Allele Frequency
MG	Myasthenia Gravis
MHC	Major Histocompatibility Complex
MS	Multiple Sclerosis
MSK	Musculoskeletal
ncRNA	Non-coding RNA
NFE	Non-Finnish Europeans
NHGRI	National Human Genome Research Institute
NIHR	National Institute for Health and Care Research

NSCLC	Non-small Cell Lung Cancer
NTC	No Template Control
OR	Odds Ratio
ORR	Objective Response Rates
PBS	Phosphate Buffered Saline
PCA	Principle Component Analysis
PCR	Polymerase Chain Reaction
PD-1	Programmed Death-1
PD-L1	Programme Death Ligand-1
PFS	Progression Free Survival
PheWAS	Phenome-wide association analysis
RA	Rheumatoid Arthritis
REF	Reference
ROC	Receiver Operating Characteristic
RPKM	Reads Per Kilobase of Exon Per Million
RT	Reverse Transcriptase
SCR	Scramble' siRNA
SD	Standard Deviation
SE	Standard Error
siRNA	Small Interfering RNA
SJS	Steven Johnson's Syndrome
SLE	Systemic Lupus Erythematosus
SNP	Single Nucleotide Polymorphism
sQTL	Splicing Quantitative Trait Loci
TBE	Tris-Borate-EDTA
TWAS	Transcriptomic-wide association analysis
T1DM	Type 1 Diabetes Mellitus

T1GS	Type 1 Diabetes Genetics
TS	Thymidylate Synthase
QC	Quality Control
UC	Ulcerative Colitis
UCSC	University of California Santa Cruz
UKBB	United Kingdom BioBank
UOB	University of Birmingham
UTR	Untranslated Region
WGS	Whole Genome Sequencing

CHAPTER 1: INTRODUCTION

1.1 Cancer treatment

Approximately 375,000 new cancer cases per year are being diagnosed in the UK, with breast, prostate, lung and bowel cancer cases comprising over half of new cancer cases (2016-2018)[2]. The incidence of cancer cases is increasing with a rise of about 4% over the last decade, indicating a growing number of patients requiring cancer treatment. Although being diagnosed with cancer has been historically associated with a poor health outcome, notable success has been achieved in cancer survival in the UK, which has doubled in the last 40 years[2]. Advances in cancer treatments have contributed considerably to this success. Cancer treatments can include localised therapy such as surgical resection and radiotherapy and systemic treatments such as chemotherapy and immunotherapy. Majority (>50%) of cancers are treated with surgery, followed by an estimated 20% treated with radiotherapy and 18% treated with chemotherapy between 2013-2020[3].

1.2 Fluoropyrimidine chemotherapy

Fluoropyrimidines including fluorouracil (5-FU) and its pro-drug capecitabine are antimetabolite chemotherapy agents widely used to treat many cancer types, especially gastrointestinal (GI) and breast cancers. 5-FU is usually being delivered intravenously as a pump or bolus infusion and capecitabine is taken orally. Both drugs can be delivered as a single agent or in combination with other systemic cancer therapies.

1.2.1 Metabolism of 5-FU chemotherapy

The liver is the main site of 5-FU metabolism but only 1-5% of the drug is converted into metabolites that will result in cytotoxic effects downstream of its pathway[4].

Figure 1.1 shows that the majority of 5-FU (~80%) is catabolised and inactivated by dihydropyrimidine dehydrogenase (DPD), the rate-limiting step of the catabolism pathway[5]. One of the main mechanisms of drug cytotoxicity is the inhibition of thymidylate synthase (TS) by metabolite FdUMP[6]. TS catalyses deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP) via reductive methylation using 5,10-methylenetetrahydrofolate (CH_2THF) as the methyl donor to generate thymidylate which is essential for DNA synthesis and repair[6]. The 5-FU metabolite FdUMP binds to TS to form a stable ternary complex with TS and CH_2THF, inhibiting the typical conversion of dUMP to dTMP. This results in a cascade of reactions leading to inhibition of DNA replication and repair and subsequent DNA damage[6]. Methylene tetrahydrofolate reductase (MTHFR) converts CH_2THF to 5-methyltetrahydrofolate, providing methyl groups that can affect DNA methylation reactions[7]. Reduced activity of MTHFR can therefore lead to increased levels of CH_2THF and potentially greater inhibition of TS, resulting in increased 5-FU cytotoxicity[8]. Apart from TS inhibition, other potential mechanisms of cytotoxicity involve the misincorporation of 5-FU metabolites FdUTP into DNA and FUTP into RNA, which leads to DNA and RNA damage respectively, resulting in cell death[4].

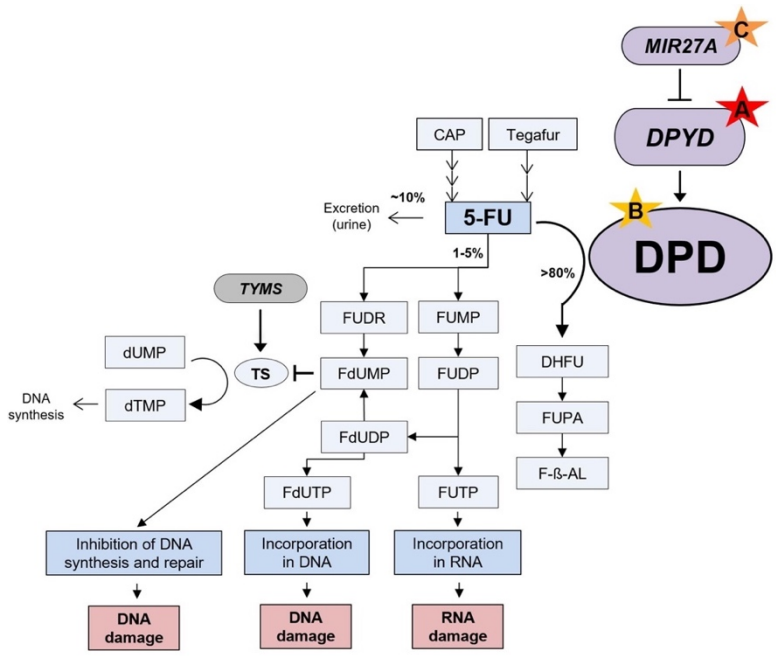


Figure 1.1 Metabolism pathway of fluoropyrimidine chemotherapy.
Oral 5-FU prodrugs Capecitabine (CAP) and Tegafur are converted to 5-FU, of which >80% of 5-FU is inactivated by the DPD enzyme (encoded by *DPYD*). 1-5% of the drug is converted into metabolites such as FduMP and FUDP, which get incorporated into DNA and RNA, causing damage and leading to cancer cell death. FdUMP inhibits the TS enzyme (encoded by *TYMS*) which is involved in DNA synthesis; this further contributes to its cytotoxic effect. Mir27-a microRNA (encoded by *MIR27A*) is involved in the post-transcriptional regulation of DPD. *DPYD* and *MIR27A* genetic polymorphisms have been linked with 5-FU-induced toxicities.
Abbreviations: 5-FU- 5-Flurouracil, CAP-capecitabine, DHFU- 5,6-dihydrofluorouracil, DPD- dihydropyrimidine dehydrogenase, DPYD- gene encoding DPD, dTMP- deoxythymidine monophosphate, dUMP- deoxyuridine monophosphate, FdUDP- fluorodeoxyuridine diphosphate, FdUMP- fluorodeoxyuridine monophosphate, FdUTP- fluorodeoxyuridine triphosphate; FUDP- fluorouridine diphosphate, FUDR- fluorodeoxyuridine, FUMP- fluorouridine monophosphate, FUPA- fluoro-β-ureidopropionate, FUTP- fluorouridine triphosphate, F-β-AL- fluoro-β-alanine; *MIR27A*- gene encoding MIR27A microRNA, TS- thymidylate synthase, TYMS- gene encoding TS. Adapted from Meulendijks et al, *Cancer Treatment Reviews*, 2016[4]

1.2.2 Activation pathway of capecitabine

Capecitabine is an orally administered fluoropyrimidine agent that is activated to 5-FU via a three-step enzymatic process involving carboxylesterase(CES), cytidine deaminase(CDA) and thymidine phosphorylase (TYMP)(Figure 1.2)[8]. Following intestinal absorption of capecitabine, the first enzymatic activation process occurs in the liver, where it is hydrolysed by CES to produce 5'-deoxy-5-fluorocytidine (5'-DFCR), of which CES1A2 and CES2 are the most potent catalysts CEA isoenzymes[8]. CDA is predominantly expressed in the liver and converts 5'-DFCR to 5'-deoxy-5-fluorouridine (5'-DFUR). Cancer cells have been observed to

have high CDA activity resulting in increased exposure to the drug and its downstream cytotoxic metabolites[9,10]. TYMP is the rate-limiting enzyme in the activation of capecitabine and converts 5'DFUR to 5-FU in the final step of the activation pathway[8]. TYMP expression is elevated in various solid tumours leading to high sensitivity of the cancer cells to the drug[11].

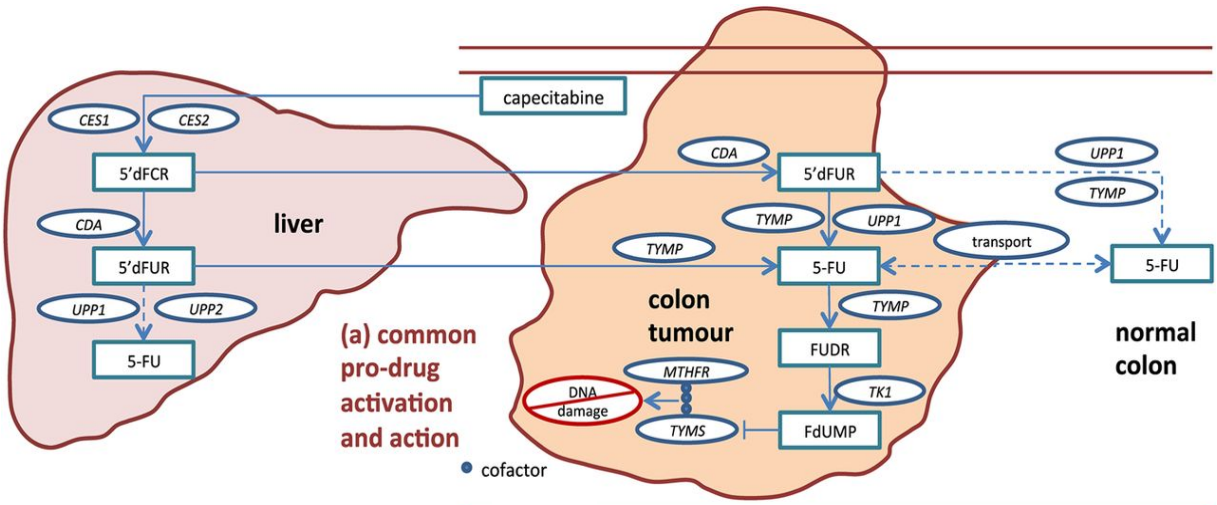


Figure 1.2 Activation and metabolism pathway of 5-FU oral prodrug capecitabine
Capecitabine is activated to 5'dFCR by enzymes CES 1 and CES2 in the liver and by CDA to 5'dFUR in the liver and colon tumour. It is then converted to 5-FU by TYMP in the tumour. MTHFR provides methyl groups that can affect DNA methylation reactions, whereby reduced MTHFR activity results in increased DNA damage and 5-FU cytotoxicity[8]. Abbreviations: CDA- cytidine deaminase, CEA- carboxylesterase, 5-FU- 5-fluorouracil, 5'dFCR- 5'-deoxy-5-flurocytidine, 5'dFUR- 5'-deoxy-5-flurouridine, FdUMP- fluorodeoxyuridine monophosphate, FUDR-fluorodeoxyuridine, MTHFR- methylene tetrahydrofolate reductase, TK1- thymidine kinase 1, TYMP- thymidine phosphorylase, UPP- uracil phosphoribosyltransferase. Adapted from: Rosmarin D et al, *Gut*, 2015[12]

1.3 The immune checkpoint inhibitors

Please see Appendix 1 for a review I wrote in 2022 (Germline genetic variation and predicting immune checkpoint inhibitor induced toxicity. Chin IS, Khan A, Olsson-Brown A, Papa S, Middleton G, Palles C. NPJ Genom Med. 2022 Dec 24;7(1):73. doi: 10.1038/s41525-022-00345-6. PMID: 36564402).

Immune checkpoints are cell surface receptors essential in maintaining normal self-tolerance to the immune system[13]. Cancer cells exploit these pathways and immune tolerance

mechanisms to avoid destruction by the immune system[14]. This exploitation happens during the immune escape phase when an increase in immune checkpoint expression in the cancer cells renders the cells resistant to the host immune system[15,16]. Immune checkpoint inhibitors (ICI) are monoclonal antibodies that predominantly act by inhibiting the action of immune checkpoints such as cytotoxic T lymphocyte antigen (CTLA-4), programmed cell death 1 (PD-1) and programmed cell death ligand 1 (PD-L1), thereby preventing the escape of cancer cells and leading to cancer cell death **(Figure 1.3)**[17].

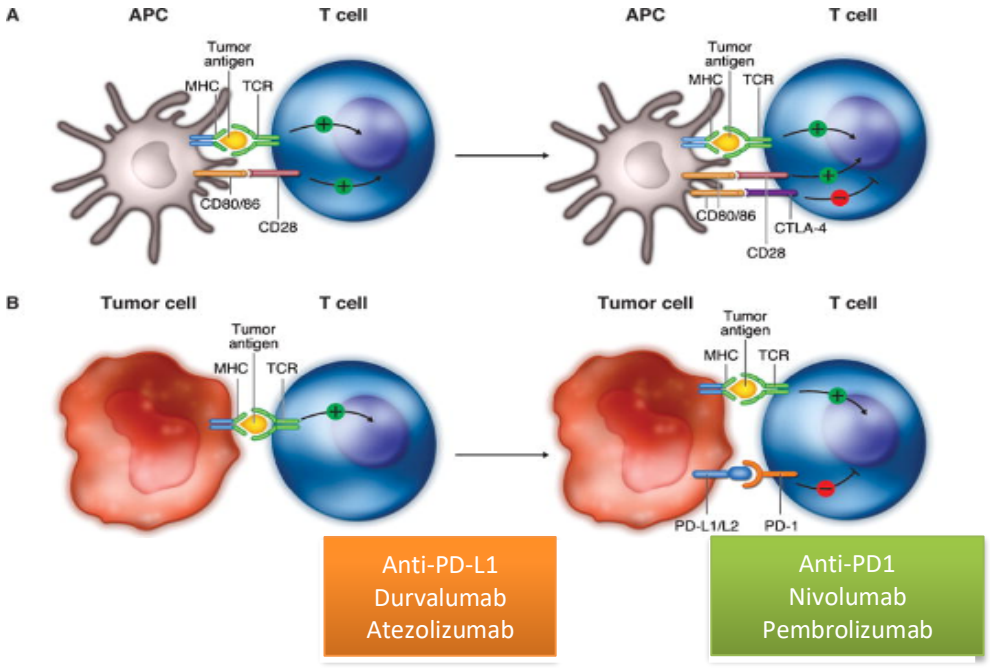


Figure 1.3 Diagram illustrating the activity of the immune checkpoint inhibitors
The presence of the tumour antigen triggers the TCR to transport CTLA-4 to the T cell surface and competes with the co-stimulatory molecule CD28 for CD80/86 expressed by the dendritic cells, creating an inhibitory signal. B. The PD1 pathway acts in peripheral tissues where upon antigen recognition, the activated T cells upregulate PD1 on their surface and bind to PD-L1 or PD-L2 on tumour cells. This interaction downregulates T-cell activity and limits the destruction of the tumour tissue. Anti-PD1/PD-L1 and anti-CTLA-4 drugs are immune checkpoint inhibitors that act on these pathways to release the inhibitory signals to reactivate T cells for cancer cell killing. Abbreviations: CTLA-4- cytotoxic T lymphocyte antigen, PD-1- programmed cell death 1, PD-L1- programmed cell death ligand 1, TCR- T cell receptor. Adapted from Carbone et al., *J Thoracic Onc*, 2015[18]

ICI was introduced as a cancer treatment in 2015 and has revolutionised the management of several cancer types, including melanoma, lung and urological cancers, which have historically had dismal prognoses in the metastatic setting. ICI therapies currently used in practice include

anti-PD1 (nivolumab, pembrolizumab), anti-PDL1 (atezolizumab, durvalumab, avelumab) and anti-CTLA-4 (ipilimumab) agents and their common treatment indications are shown in **Table 1.1**.

Table 1.1 Types of ICI therapies and their therapeutic indications

Class of ICI agent	ICI therapy	Cancer type indication
Anti-PD-1	Pembrolizumab	Lung, melanoma, urothelial, head and neck cancer, Hodgkin's disease
	Nivolumab	Melanoma, lung, renal, head and neck, urothelial cancer, Hodgkin's disease
	Cemiplimab	Squamous cell carcinoma
Anti-PD-L1	Durvalumab	Lung cancer
	Atezolizumab	Lung, urothelial cancer, breast
	Avelumab	Merkel cell and renal cancer
Anti-CTLA-4	Ipilimumab	Melanoma

The current ICI agents used in practice have multiple therapeutic indications across different cancer types. Abbreviations: CTLA-4- cytotoxic T lymphocyte antigen, ICI- immune checkpoint inhibitor, PD-1- programmed cell death 1, PD-L1- programmed cell death ligand 1

1.4 Toxicities related to cancer treatment

Despite the high efficacy of fluoropyrimidine and ICI as cancer therapies, management of the associated treatment-related adverse events remain a crucial clinical challenge. These toxic effects are often variable and unpredictable and can cause considerable morbidity, life-threatening complications and adverse impact on long-term health and quality of life. Adverse events can also lead to increased hospital admissions and healthcare system costs[19]. The increasing cancer incidence indicate a growing number of patients requiring cancer treatment leading to more patients presenting with treatment-related adverse events.

1.4.1 Fluoropyrimidine-induced toxicity

Toxicities associated with fluoropyrimidine-based chemotherapy include diarrhoea, neutropenia and hand-foot syndrome (HFS), which can be challenging to manage in clinical practice. Approximately a third of patients with colorectal cancer experienced ≥grade 3 Common Terminology Criteria for Adverse Events (CTCAE) grade 3 or higher toxicities from 5-FU or capecitabine[4,20]. Up to 20% of patients are admitted to hospitals due to treatment-related toxicities, with a fatality rate of approximately 1%[4].

1.4.2 Cardiovascular toxicity

Cardiovascular toxicity related to 5-FU is relatively uncommon but can be severe and life-threatening. A variable incidence of cases from 0.3% up to 68% has been reported, while the frequency of events related to oral prodrug capecitabine ranged from 2.8-34.6%[21]. Clinical manifestations of these cardiac events include angina, myocardial infarction, arrhythmias, cardiogenic shock, heart failure or sudden death[22]. Mortality rates reported from these events ranged from 2.2-13.3%[23]. The exact pathophysiology and mechanisms behind the development of cardiotoxicity are still unestablished. It is hypothesized to be multifactorial, of which abnormal coronary vasospasms, endothelial damage, increased thrombogenicity, and direct myocardial injury have been suggested as potential mechanisms[21,22]. Proposed risk factors leading to cardiovascular toxicity include cardiovascular risk factors, cardiac comorbidities, combination therapies with chemotherapy, radiotherapy or targeted therapy, and inheritance of genetic polymorphisms. A case study of a patient with pancreatic cancer who developed 5-FU and capecitabine-induced cardiovascular toxicity was found to be a heterozygous carrier of the *DPYD* P453L (1358C>T) variant[24]. However, there is contradicting evidence behind the associations of these risk factors, with a lack of studies to

date assessing the relationship between genetic polymorphisms and cardiotoxicity. More research is required to establish these relationships to help understand the pathogenesis behind this potentially fatal adverse event and identify biomarkers to detect patients at risk in clinical practice.

1.4.2.1 Pre-clinical studies assessing effects of 5-FU on cardiac cells

Given the uncertainty surrounding the pathogenesis of 5-FU-induced cardiotoxicity, several pre-clinical studies have been conducted to determine the likely causal mechanism behind this unpredictable toxicity.

Lamberti et al. studied the effects of 5-FU (+/- levofolene) in H9c2 rat cardiomyocytes and HT-29 human colon adenocarcinoma cell lines using an increasing drug concentration for 24, 48 and 72 hours. The study found that 5-FU led to a dose- and time-dependent inhibition of cell growth in both cell lines, as assessed by the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. As assessed by flow cytometry, apoptosis was observed in the H9c2 cell line, enhanced by levofolene[25]. 5-FU induced apoptosis observed in rat cardiomyocytes was accompanied by increased activation of the caspase cascade and inhibition of apoptosis by reactive oxygen species (ROS) scavenger N-acetyl cysteine (NAC), suggesting that 5-FU induced apoptosis in these cells is likely mitochondria-mediated and due to the increase of ROS intracellularly[25]. A study by Focaccetti et al. analysed the effects of 5-FU on human cardiomyocytes and endothelial cell lines and found features of autophagy, dose-dependent increase in ROS production and senescence induced in both cell types following treatment with 5-FU[26].

1.5 ICI induced toxicity

Despite the notable successes of ICI therapy in improving patients' survival, toxicities associated with ICI, also known as immune-related adverse events (irAE), are potentially fatal and life-changing. As the indication for using ICI continues to expand and move earlier in the treatment pathway in the adjuvant and neoadjuvant setting, identifying patients at risk of developing toxicities is important to minimise toxicity risk while maximising the treatment benefit gained by patients.

1.5.1 The spectrum and challenges of immune-related adverse events

IrAEs are deemed to be caused by an over-active immune response and disruption in immune homeostasis, manifesting as autoimmune-type events[17]. irAEs are unpredictable in onset, severity and clinical presentation and can affect multiple organ systems. There is a wide range of clinical manifestations from irAE, including cutaneous (manifesting as rash, pruritus, vitiligo), endocrine (thyroid, pituitary, adrenal disorders), gastrointestinal (colitis), lung (pneumonitis), musculoskeletal (arthralgia, myalgia) and liver (hepatitis) toxicities

1.5.2 Toxicity profile of different ICI therapies

The incidence and severity of irAEs vary according to the ICI treatment regime. The variation in the incidence and spectrum of irAE between different ICI therapies can also be due to their mechanism of action at different parts of the cancer immunity cycle. Anti-CTLA-4 agents predominantly act during the initial T cell priming stage at the lymphoid tissue, while anti-PD1 or anti-PD-L1 agents work in the peripheral tissues to stimulate the activity of T cells[27]. **Table 1.4** shows the frequency of different side effects by treatment type PD-L1/PD1 and CTLA-

4[28–30]. Gastrointestinal and cutaneous toxicities are the most common irAEs, in up to 65% of patients treated with an ICI. Rarer toxicities such as type I diabetes (T1DM), adrenal insufficiency, and ocular toxicities can also occur and tend to affect ~1 % of patients.

Table 1.2 Frequency of different irAE according to the ICI treatment type

ICI agents	Anti-PD-1		Anti-PD-L1		Anti-CTLA-4		Anti-CTLA-4 and anti-PD-1	
irAE	**Max frequency of any grade events (%)**	**Max frequency of ≥ Grade 3 events (%)**	**Max frequency of any grade events (%)**	**Max frequency of ≥ Grade 3 events (%)**	**Max frequency of any grade events (%)**	**Max frequency of ≥ grade 3 events (%)**	**Max frequency of any grade events (%)**	**Max frequency of ≥ Grade 3 events (%)**
Pneumonitis	2.4-3.6%	0.7%	3.1%	1.1%	0.40%	0.1%	5.3%-7.8%	0.2-1.1%
Cutaneous	4-28.5%	0-1.3%	<0.1-16%	0-0.6%	1.1-25.9%	0-1.3%	37.7-65%	0.3-7.6%
Diarrhoea	12.1-13%	1-1.5%	7.3%-16.3%	0-0.6%	27%	5-7.4%	22.3%-46.7%	0-15.8%
Colitis	0.7%-2%	0.4%-2%	0.9%	0.3%	5.7- 8%	4.1%-5%	46%	16%
Hepatitis	0.1-7%	0%-1.5%	0.4.%	0-0.4%	0-2.6%	0-0.9%	13.4-29.5%	1.1-15.4%
Thyroiditis	1.6%-9.7%	0.1%	0-2.2%	<0.1%	0.5- 2%	<0.1%	24-27.2%	0.3-1.3%
Pituitary	0.2%	0.1%	<0.1%	<0.1%	1.2-4%	0.80%	1.4-4%	0.4-2.4%
Adrenal	0.6%	0.2%	0.70%	0.4%	0.40%	0.20%	0.3-3.5%	0.2-1.4%
Diabetes	0.1%	0.1%	<0.1%	<0.1%	<0.1%	-	0.5%	0.5%
Neurological	6%	2%	<0.1%	<0.1%	3.8-4.5%	1.9%	12%	-
Renal	2.7%	<0.1-0.4%	0.3%	<0.1%	<1%	-	5.1-8.6%	0.5-1.7%
Rheumatological	0.1-43%	0-0.2%	0-6.2%	<0.1%	0-2.1%%	0-0.2%	5-14%	<1-1%
Ocular	0.1-1%	-	<1%	-	<1%	-	-	-

The incidence of irAE varies according to the type of organ-specific irAE and ICI treatment received. Abbreviations: CTLA-4- cytotoxic T lymphocyte antigen, ICI- immune checkpoint inhibitor, PD-1- programmed cell death 1, PD-L1- programmed cell death ligand 1. Source data: Khoja et al, *Ann Oncol* 2017[28], Haanen et al, *Ann. Oncol.* 2017[29], Electronic Medicines Compendium[30]

A higher incidence of irAEs has been reported in patients receiving combination ICI compared to single-agent ICI therapies[31]. A randomised phase 3 trial showed that 95.5% of melanoma patients on combined nivolumab and ipilimumab experienced treatment-related toxicities of any grade, and 55% of patients developed ≥grade 3 toxicities[31]. High rates of ≥grade 3 toxicities (67.2%) are also observed in lung cancer patients treated with concurrent pembrolizumab and chemotherapy[32]. The incidence of irAEs following single-agent ICI treatment is lower. Approximately a third of patients treated with anti-CTLA-4 alone experienced ≥grade 3 events, while a meta-analysis found 14% of patients on anti-PD1 or anti-PD-L1 therapy experienced ≥grade 3 events[28,29,33]. Anti-PD-L1 treatment generally leads to fewer irAEs than anti-PD1, possibly due to a lack of inhibition of the PD-L2 ligand, which is involved in T cell regulation[34].

A systematic review of 6,938 patients found colitis, hypophysitis and skin toxicities more common with anti-CTLA-4 therapy, while pneumonitis, hypothyroidism, vitiligo and musculoskeletal toxicities occurred more with anti-PD1 treatment[28]. Ectopic expression of CTLA-4 in the pituitary endocrine cells causing activation of the complement and inflammatory pathways could explain why immune-related hypophysitis appeared more common with anti-CTLA-4 therapy[35]. A meta-analysis of 112 trials reported fatality rates related to irAE ranging from 0.36-1.23% in patients receiving monotherapy and combination ICI treatment respectively[36]. Colitis and myocarditis were frequent causes of death in patients receiving combined ICI therapy, whereas pneumonitis, hepatitis and neuro-toxicities most commonly contributed to anti-PD-1 or PD-L1 related fatalities[36].

1.5.3 Patient and tumour characteristics that may impact on risk of irAEs

Chennamadhavuni et al. have reviewed the patient and tumour characteristics associated with the risk of irAEs[37]. It is difficult to make general statements about individual risk factors as their influence is context dependent, varying by type of ICI used and organ at risk of an irAEs. There is general agreement across studies that endocrine toxicities and pneumonitis are more common in younger patients, while skin toxicities are more frequent in older patients[38,39]. There are also reports that sex influences risk, with thyroid irAEs being more common in women while neurological, dermal and vascular events are more common in men[40]. Higher body mass index (BMI) and performance status have also been associated with an increased risk of irAE[41,42]. Specific co-morbidities have been shown to increase the risk of certain irAEs, e.g. interstitial lung disease is associated with an increased risk of pneumonitis[43].

Studies involving patients treated with anti-PD1 have found that melanoma patients experienced more gastrointestinal and skin irAEs, while pneumonitis was more common in lung and renal cancer patients[28]. When considering the incidence of any grade irAE or ≥grade 3 irAEs, Wang et al. found no significant difference in incidence by cancer type; however, melanoma patients did experience the highest incidence of any grade irAEs[33].

1.5.4 Time to onset of irAE

The timing of when irAEs occur varies according to the toxicity and treatment type. In an analysis of 8436 patients treated with ICI therapies, the pooled median time to onset of any toxicity of any grade ranged from 2.2 to 14.8 weeks[44]. Skin, infusion reaction and gastrointestinal events had the shortest median onset time (2.2-8.9 weeks), whereas renal

toxicities took the longest to manifest (14.8 weeks)[44]. Late-onset toxicities have also been reported up to almost one year after cessation of ICI therapy[29]. When considering ≥grade 3 events, the median onset time was 7 weeks for anti-CTLA-4 treatments, 27.5 weeks for anti-PD1/PD-L1 therapy and 7.9 weeks for patients treated with doublet ICI therapy[44].

Some irAE can result in persistent and long-term effects. The median time to resolution of irAE can range from 0.1 weeks for hypersensitivity/infusion reactions to 54.3 weeks for endocrine irAE[44]. The lasting impact of ICI was demonstrated objectively in a study of patients with anti-CTLA-4 induced enterocolitis, where evidence of colitis was still observed on endoscopy at a median of 4 months from the onset of symptoms[45]. Other persistent irAEs that can continue beyond completion of treatment include ICI-induced T1DM and inflammatory arthritis. These toxicities cause considerable morbidity, reliance on lifelong medication and expensive costs to healthcare systems[45–47]. Steroids are the primary immunosuppressive treatment to counteract irAE and can potentially cause debilitating effects on bone health, adrenal insufficiency and steroid-induced diabetes[48]. Immunomodulatory therapies may also be required to treat steroid-refractory or recurrent irAE that can cause profound immunosuppression. The significant impact of irAE on patients' long-term health and quality of life needs to be considered alongside the potentially longer survival benefit gained from ICI therapy.

1.6 Role of germline genetic variation in determining toxicity risk

There is a currently a lack of established predictive tools or biomarkers can help predict the risk of developing treatment-related adverse events. For patients receiving ICI therapies, clinicians mainly rely on the patient's past clinical history, including the presence of

autoimmune disease (AD), to estimate risk. However, as patients with ADs tend to be excluded from most clinical trials of ICIs, it is hard to determine the effectiveness of this strategy.

The ability to predict patients who are at risk of adverse events will be clinically helpful to better manage their cancer whilst developing strategies to prevent toxicity occurrence.

Inherited genetic variation has been shown to influence the development of treatment-related toxicity[49]. Associations between germline genetic polymorphisms and toxicity have already been established for several chemotherapy agents[50]. For example, variants in the dihydropyrimidine dehydrogenase *(DPYD)* gene with statistical support of an association with toxicity are being screened for before administration of fluoropyrimidine chemotherapy due to the risk of low-activity variants leading to ≥grade 3 toxicities[50]. Another example involves inheriting a dysfunctional variant of *UDP glucuronosyltransferase 1A1 (UGT1A1),* which can increase exposure and toxicity from irinotecan chemotherapy[50]. Variants of the gene encoding thiopurine S-methyltransferase (TPMT) was also found to be associated severe myelosuppression in haematological cancer patients receiving thiopurine drugs[51]. However, these known toxicity risk variants only account for a small fraction of toxicity induced by these agents.

We hypothesise that additional genetic variants could predict those at risk of fluoropyrimidine toxicity and that an inherited genetic predisposition also drives immunotherapy toxicity. The ability to predict patients likely to experience toxicity when treated with particular drugs could allow treatment dose or regime modifications. Mapping toxicity-associated variants to functional variants altering gene expression and function will help to understand the mechanisms leading to toxicity.

1.6.1 Germline determinants of host immunity

Host genetic factors are hypothesized to play an essential role in ICI-related outcomes, given that they are strong determinants of immune homeostasis and our immunological status[52]. In the case of ICI therapy, the activity of the host's immune system can strongly influence treatment response. Inherited genetic factors may explain the variation in treatment outcomes seen among patients receiving similar cancer treatments and the apparent disparity in immune responses between these patients[53].

1.6.1.1 The cancer immunity cycle

The cancer immunity cycle consists of stepwise events delineating the anti-cancer immune response. **Figure 1.4** illustrates the seven steps in the cycle, which is characterised by stimulatory and inhibitory immune factors that regulate the immune feedback mechanisms to activate the immune response but prevent autoimmunity[27].

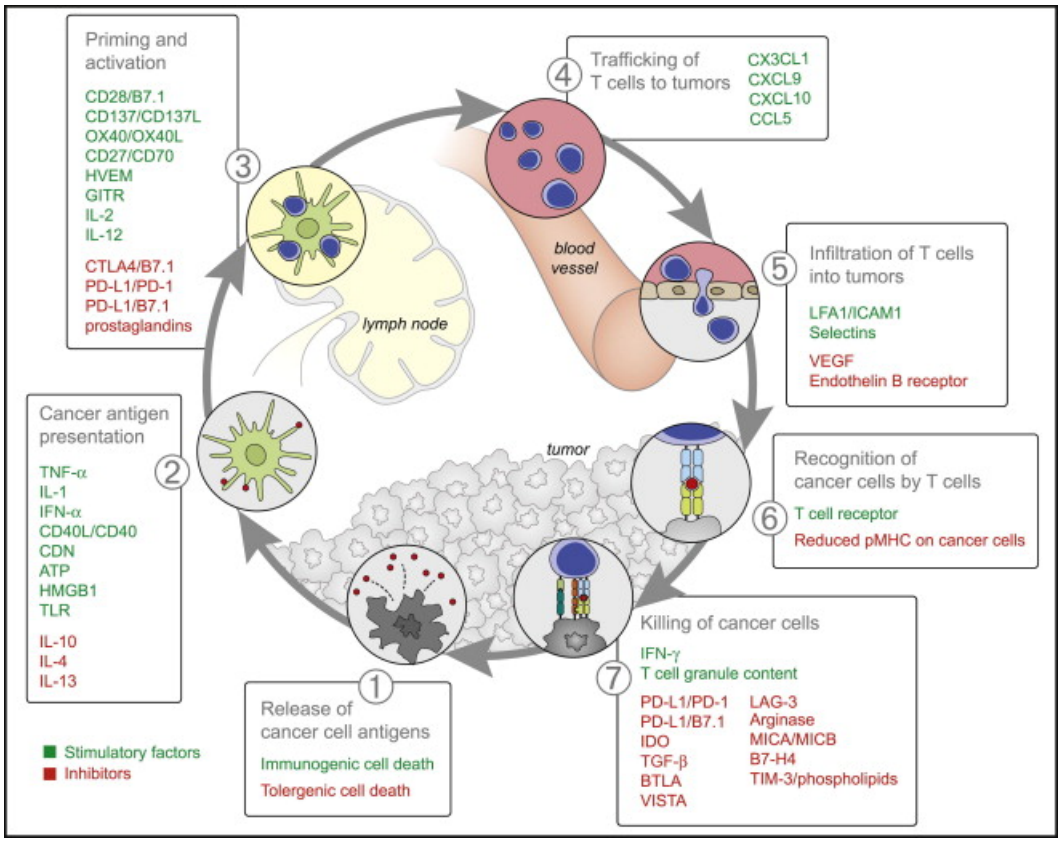


Figure 1.4 Cancer immunity cycle

The cycle involved stepwise events characterising the immune response against cancer cells. Different stimulatory and inhibitory immune factors play a role in each step by activating or suppressing the immune response towards cancer cells. CTLA4, PD-1, PD-L1 and LAG-3 are immune checkpoints that inhibit T cell activation and prevent the killing of cancer cells. Adapted from Chen et al., *Immunity*, 2013[27]

1.6.1.2 Interplay between host genetics and immuno-modulatory cells and molecules

Host genetics can control the activity of numerous immune cell types[52]. Many immune traits are heritable. Those with the highest heritability include CD39 on CD4 T cells functioning as regulatory T cells and CD32 expression on dendritic cells[54]. Dendritic cells have the largest proportion of highly heritable traits, followed by CD4 T cells and CD8 cells[55]. Hundreds of SNPs have been identified to be associated with the abundance of blood cells, immune cell marker expression and levels of molecules such as growth factors, cytokines, and MHC-related proteins. **Table 1.5** shows the SNPs identified from the GWAS catalogue that have been significantly associated with immune cell traits involved in the cancer immunity cycle at genome-wide significance ($P \leq 5x10^{-8}$)[56].

Table 1.3 Genetic variants associated with immuno-modulatory traits in the cancer immunity cycle

Immune traits	No. of SNPs reaching GWAS significance ($P \leq 5x10^{-8}$) and in linkage equilibrium	Average effect sizes (beta) of alleles	Largest discovery cohort's sample size (ethnic population)	Largest replication cohort sample size (ethnic population)
TNFα	1	2.13	13, 577 (Finnish)	NA
IL-1α	1	0.74	790 (Hispanic, European, Asian, South Asian or African American)	NA
IL-1β	1	0.33	997 (European)	338 (Arab, South Asian and Filipino)
IFNα	1	1.80	410 (European)	1165 (European, African American)
CD40	2	0.89	982 (Scottish)	NA
CD40L	1	0.96	653 (European)	317
IL-10	8	0.08	764 (Hispanic, European, Asian, South Asian or African American)	425 (West African)
IL-4	1	0.40	764 (Hispanic, European, Asian, South Asian or African American)	NA
IL-13	1	0.42	3,557 (Finnish)	NA
CD27/CD70[a]	2	0.64	997 (European)	338 (Arab, South Asian and Filipino)
IL-2	1	0.27	764 (Hispanic, European, Asian, South Asian or African American)	NA
IL-12[a]	5	0.21	9,263 (Finnish)	NA
CCL5	1	0.33	3,421 (Finnish)	NA
LFA1/ICAM1[a]	6	3.47	22,435 (European)	9,813
VEGF[a]	18	0.17	13,577 (Finnish)	2800 (European, Sorbian)
IFNγ	2	0.31	7,701 (Finnish)	NA
MHC class I polypeptide-related sequence A (MICA)[a]	27	0.77	997 (European)	338 (Arab, South Asian and Filipino)
MHC class I polypeptide-related sequence B (MICB)	8	0.63	997 (European)	338 (Arab, South Asian and Filipino)

Variants that reached GWAS significance ($P \leq 5x10^{-8}$) were identified from the GWAS catalogue from March-April 2021. For a SNP to be included in this table, the following summary statistics had to be available: effect allele, non-effect allele, beta and P-value (minimal statistics required for including a variant in a polygenic risk score). Linkage disequilibrium (LD) between the lead SNP and any other SNPs in the region had to be <0.8. The average absolute beta value of all SNPS is reported for traits with >1 correlated SNP. SNPs with effect sizes reported in units measuring gene/protein expression change were excluded. Abbreviations: GWAS- genome-wide association study, NA-not available, SNPs- single nucleotide polymorphism. [a] Some SNPs for this trait have beta effect sizes that are inverse transformed and/or scaled to standard deviation units

Immune checkpoints CTLA-4, PD-1 and PD-L1 targeted by ICI therapies elicit an inhibitory effect in the priming and cancer-cell-killing phases of the cancer immunity cycle[27]. *CTLA-4* gene polymorphisms are associated with Graves' disease, hypothyroidism and T1DM[57]. The most likely causal polymorphisms in a non-coding 3'UTR region affect *CTLA-4* expression and are presumed to increase T-cell reactivity[57]. A missense variant, *CTLA-4* Y60C, was also linked with early-onset Crohn's disease[58].

Vogelsang et al. identified 385 expression quantitative trait loci (eQTL) SNPs associated with the expression of 268 genes involved in immunoregulation[59]. By testing the 40 most significant eQTLs in >1000 melanoma patients, they identified two SNPs, rs6673928 (with influence on *IL-19* expression) and rs6695772 (with influence on *BATF3* expression), that were significantly associated with overall survival (OS) and both mapped to chromosome 1q32[59]. As eQTLs are associated with survival and predict prognosis from an immunogenic malignancy like melanoma, it would be interesting to understand if such variants also predict toxicity.

1.6.2 Genetics of Autoimmune Disease

The close similarity in clinical presentation between patients with irAE and autoimmune disease (AD) justifies further investigations into whether this resemblance is explained by shared genetic susceptibility. Candidate genes known to be correlated with AD are strong candidates in the search for potential genetic determinants of irAE.

Hundreds of disease risk loci marked by thousands of SNPs have been identified as associated with >60 AD in genome-wide association studies (GWAS)[52]. Each risk loci is marked by one or more independent polymorphisms carrying small to moderate effect sizes[60]. The loci map to non-coding regions enriched in distant regulatory elements[60]. Table 3 shows the number

of genetic loci in linkage equilibrium that reached GWAS significance for common autoimmune disease traits from the GWAS catalogue from March to April 2021[56].

Table 1.4 Genetic variants associated with common autoimmune disease traits

Autoimmune disease trait	No. of SNPs reaching GWAS significance ($P \leq 5x10^{-8}$) and in linkage equilibrium	Average effect sizes (OR) of alleles	Largest discovery cohort's sample size (ethnic population)	Largest replication cohort sample size (ethnic population)
Rheumatoid arthritis	186	1.16	22628 cases, 288664 cases (European, East Asian)	24107 cases, 79295 controls (East Asian, European)
Psoriasis	127	1.75	11,988 cases, 275,335 controls (European)	11301 cases, 19879 controls (European)
Crohn's disease	180	1.23	19,085 cases, 34,213 controls (European)	16619 cases, 31766 controls (European, Iranian, Indian, East Asian)
Ulcerative colitis	80	1.32	12,924 cases, 21,442 controls (European)	25,683 cases, 17,015 controls (European)
Systemic lupus erythematosus	184	1.41	11590 cases, 15984 controls (European, African American, Hispanic)	1,387 cases, 28,564 controls (Japanese)
Ankylosing spondylitis	55	3.74	8,726 cases, 34,213 controls (European)	2,111 cases, 4,483 controls (European)
Autoimmune thyroid disease	141	1.48	30,234 cases, 724,172 controls (European)	7,891 cases, 8,351 controls (Han Chinese)
Coeliac disease	59	1.29	11,489 cases, 22,308 controls (European)	12041 cases, 12228 controls (European, Indian)
Type 1 diabetes	66	1.30	9,934 cases, 16,956 controls (European)	4,329 cases, 9,543 controls (European)
Vitiligo	70	1.37	2,853 cases, 37,405 controls (European)	6,623 cases, 10,740 controls (East Asian)
Addison's disease	10	2.23	1,223 cases, 4,097 controls (European)	NA
Myasthenia gravis	8	1.68	972 cases, 1,977 controls (European)	423 cases, 467 controls (European)
Steven Johnson's Syndrome	5	2.53	424 cases, 1,881 controls (European)	102 cases, 469 controls (East Asian, Indian, Brazilian)
Autoimmune hepatitis	1	2.90	649 cases, 13,436 controls (European)	451 cases, 4103 controls (European)

Variants that reached GWAS significance ($P \leq 5x10^{-8}$) were identified from the GWAS catalogue from March-April 2021. For a SNP to be included in this summary of variants, the following summary statistics had to be available: effect allele, non-effect allele, beta (or odds ratio) and p-value (minimal statistics required for including a variant in a polygenic risk score). LD between the lead SNP in a region and any other SNPs in the region had to be <0.8. The average absolute OR of all SNPs associated with the relevant trait is reported. Abbreviations: GWAS-genome-wide association study, OR- odds ratio, NA-data not available, SNPs- single nucleotide polymorphism.

1.6.2.1 Association between the MHC locus and autoimmune disease

The major histocompatibility complex (MHC) locus is a 3.6Mb genomic region comprising more than 250 genes commonly linked with AD[60]. In particular, the human leukocyte antigen (*HLA*) gene plays a vital role towards AD occurrence as it encodes for the receptors expressed by antigen-presenting cells[60]. Therefore, variation in its genotype could influence the immune response and the risk of developing an autoimmune condition. *HLA* genes are often challenging to map, given their polymorphic nature and the long-range linkage disequilibrium observed across the MHC region[61]. *HLA* class I alleles tend to be associated with seronegative and specific types of AD[61]. For example, *HLA-B27* and *HLA-A*0201* are linked with ankylosing spondylitis, *HLA-Cw6* with psoriasis, and *HLA-B51* with Behcet's disease[61]. *HLA* class II alleles are typically associated with seropositive AD and involve the *HLA-D* region[62]. Examples include the linkage between the *HLA-DR3-DQ2* haplotype with classical autoimmune phenotypes such as coeliac disease, T1DM, systemic lupus erythematosus (SLE) and autoimmune thyroid disease, as well as *HLA-DRB1* with rheumatoid arthritis (RA)[61]. Detailed SNP typing and imputation have enabled the linkage of individual AD to multiple independent MHC genes[61]. For example, imputation of classical *HLA* alleles and amino-acid polymorphisms further refined the 'shared epitope' *HLA-DRB1* alleles linked with RA to three amino acids in the peptide-binding groove as well as identifying associations at *HLA-B* and *HLA-DPβ1*[63]. Other AD associated with multiple *HLA* alleles include psoriasis, ankylosing spondylitis, multiple sclerosis and T1DM[61]. The complexity in AD genetics is also demonstrated in T1DM, where specific combinations of *HLA DRB1, DQA1 and DQB1* alleles are associated with either risk or protection from disease[64].

Research on *HLA* variation in pharmacogenetics is not novel, and some alleles are currently tested for in clinical practice. For example, individuals starting Abacavir should be screened

for *HLA-B*57:01* as this is associated with severe hypersensitivity reaction to this drug[65]. Stevens-Johnsons Syndrome is also a rare but severe drug hypersensitivity reaction induced by carbamazepine influenced by *HLA* variation[66].

1.6.2.2 Non-MHC susceptibility loci

Non-*HLA* genes at the MHC can also play a role in the inheritance of AD, and several hundred associations with one or more AD have been reported to date[62]. These loci are usually within or close to immune-related genes involved in modulating our immune response with roles in antigen presentation, recognition, signalling and immune regulation expressed by dendritic cells, T cells and B cells[62]. Some of these genes are important in the cells targeted by ICIs with roles in the priming and activation phase of the cancer immunity cycle. Variations in these genes are candidates for explaining the risk of irAEs. Examples include genetic variants in the IL-23 and Th1 helper pathways, which are involved in the immune regulation of exogenous antigens and have been implicated in immune-mediated and AD[61] and the interferon regulatory factor (IRF) proteins, which are transcriptions factors involved in regulation of the immune response[61].

1.6.2.3 Challenges in genetic studies of autoimmune disease will also apply to studies of irAE susceptibility

AD genetics is complicated by the polygenic nature of the disease, where multiple genetic variants explain disease risk, and by genetic pleiotropy, defined as the association of one genetic locus with multiple different AD[52]. For example, *PTPN22* 1858C>T predisposes towards T1DM but is protective against Crohn's disease[67,68]. Genetic determinants of susceptibility can also vary between different ethnic populations. This variation is demonstrated in studies involving the Han Chinese population, which have found strong associations of the *HLA-B*1502* locus with a rare immune-mediated cutaneous disease Steven

Johnson's Syndrome and toxic epidermal necrolysis induced by carbamazepine[69]. However, the association was not detected in European populations[70]. Pharmacoethnicity is also a known problem evidenced by the *DPYD* risk variants currently recommended for testing by CPIC to identify patients at risk of ≥grade 3 5-FU toxicity. The variants recommended for testing have much lower frequencies in Asian and African populations than in European populations[71,72].

1.6.2.4 ICI in patients with autoimmune disease

Patients with a history of AD are deemed to be at a higher risk of developing irAE when treated with ICI therapy. The concern is that a dysregulated immune response caused by autoreactive T-cells against host tissue might potentially enhance existing autoimmunity[73]. A pre-clinical study found that mice deficient in CTLA-4 suffered early lethality from diffuse infiltration of T-cell blasts involving multiple organs[74]. Associations between polymorphisms of target cell genes *PD-1, PD-L1 and CTLA-4* and a range of ADs have also been reported[75].

Safety data in patients with AD treated with ICI is limited as patients with active AD requiring recent immunosuppressive therapy were excluded from trials studying the efficacy of ICI. Having concurrent AD is not uncommon in cancer patients. Epidemiological studies from the US have revealed that 13.5% of lung cancer patients also had an AD[76]. Menzies et al. evaluated 119 patients treated with anti-PD-1 and found 38% of patients with AD had an exacerbation of their AD requiring immunosuppression, but these were mainly mild events[77]. 29% of patients in the study with an AD developed irAEs with 10% experiencing ≥grade 3 toxicities, comparable with other clinical trials[77]. However, the safety profile may differ with anti-CTLA-4 therapy; one study found 33% of patients with a pre-existing AD developed grade 3 to 5 toxicities on ipilimumab, indicating that patients receiving anti-CTLA-4 may require more stringent monitoring[78].

1.7 Identification of genetic factors predisposing to toxicity

1.7.1 Methodology for identifying genetic variants predictive of toxicity

Selecting the optimal methodology to identify candidate variants and genes predisposing towards cancer treatment-induced toxicity is paramount towards developing a reliable predictive risk model of toxicity.

1.7.1.1 Genome-wide association studies

One established method of determining novel genetic factors predisposing to certain traits and diseases is conducting genome-wide association studies (GWAS). GWAS are hypothesis-free methods to identify the associations between genetic loci and phenotypic traits[79]. GWAS findings have helped researchers understand disease susceptibility, discover new drug targets and biomarkers, and personalise therapeutic options[80]. GWAS has begun to be applied to the detection of common variants predisposing to treatment-induced adverse effects of both traditional chemotherapies and immunotherapies. Most studies use single nucleotide polymorphism (SNP) array technology combined with imputation. Still, the use of off-target reads from gene panel sequencing has inventively been used to perform the first GWAS of immune checkpoint inhibitor (ICI) induced immune-related adverse events (irAEs)[80–82]. Whole genome sequencing (WGS) with standard depth (30x) is a more comprehensive genotyping modality and enables rare variants to be accurately assessed. However, it is much more expensive than SNP array genotyping in terms of experimental cost and analytical complexity[83]. Although whole exome sequencing is potentially more cost-effective than WGS, the non-coding regulatory regions are not assessed with this approach which could be functionally important[83].

Whilst GWAS do not generally identify causal variants, numerous methods assist with identifying the most likely causal variants underlying a GWAS signal[84] and in linkage disequilibrium with the lead genotyped or imputed SNPs that may only be tags. Definitively identifying the gene regulated by a GWAS signal is not a trivial task. Still, much progress has been made thanks to combining GWAS data with extensive information about tissue-specific expression and cell type-specific chromatin interactions[84].

The GWAS approach does have some limitations. These include the need to overcome the multiple testing burden, the inability to explain the missing heritability of complex traits and the generation of signals related to population stratification[80]. Most limitations can be overcome by increasing the sample size, using denser genome-wide SNP arrays or using low pass (0.5x) whole genome sequencing instead of arrays. Low-pass sequencing has better genomic coverage but higher computational costs[80].

Most GWAS of complex diseases and human traits consists of over 1000 cases and 1000 controls and detect common variants with modest individual effect sizes (Odds ratio (OR)< 1.2)[80]. Although GWAS require large sample sizes, pharmacogenetics studies have reported larger per allele effect sizes than those detected in studies of complex diseases[85]. Studies of ~500 patients with toxicity data will therefore be powered to pick up large effect sizes above an OR of 2 (assuming a grade 3 toxicity incidence of ~10%). Studies with large sample sizes will be needed to detect the full spectrum of common and rare variants explaining an individual's risk of developing drug-induced toxicities.

1.7.1.2 Polygenic risk score analysis

Despite the promising translative utility of GWAS, the genetic architecture of most common adult-onset diseases is complex[86]. Genetic studies of most diseases have found that common variants of small effect sizes accounted for their heritability with minimal impact from rare risk variants with moderate or large effect sizes[86]. Therefore, polygenic risk profiling using polygenic risk scores (PRS) is a valuable adjunct to GWAS. PRS is a measure of an individual's genetic susceptibility to a disease trait. It is generated by the weighted sum of the risk alleles they carry using the summary statistics from pre-existing GWAS data[87]. Genetic variants identified by GWAS studies of treatment-induced toxicity could be used to generate PRS to identify those at high or low risk of toxicity[86]. To test the predictive ability of a PRS model in identifying patients at risk, an area under the receiving operative curve (AUC-ROC) analysis can be conducted based on the PRS scores generated for each individual. The AUC defines the probability that an individual with toxicity will have a higher risk than those without toxicity[88]. PRS can be combined with clinical risk factors to generate predictive models that can potentially be clinically helpful in stratifying patients and identifying high-risk patients who warrant additional surveillance or intervention.

Several limitations of PRS need to be considered when applying to clinical predictive models. PRS will only be valid if there is a strong genetic component/heritability and if a large amount of the variation in risk due to genetics has been uncovered through existing GWAS studies. PRS will only be generalisable to multi-ethnic populations if the variants included are risk variants in multiple populations[86].

1.7.2 Genetic factors predisposing to fluoropyrimidine toxicities

Several factors can influence the development of toxicities, including the dosing schedule, duration, type of treatment regime and liver function. Genetic polymorphisms have been increasingly studied regarding their role in influencing the development of chemotherapy toxicity, given their potential to affect the expression levels of drug-metabolising enzymes, transporters and drug targets[89].

1.7.2.1 DPYD gene polymorphisms in 5-FU toxicity screening

Up to 8% of the population have deficient DPD activity, and approximately 0.01-0.1% have complete DPD deficiency, leading to severe and potentially life-threatening 5-FU toxicity[4,90]. DPD deficiency has been associated with severe ≥grade 3 5-FU toxicity, where up to 80% of individuals deficient in DPD activity can experience severe toxicities compared to about 20% who are not DPD deficient[89]. In particular, neutropenia commonly occurs in individuals with DPD deficiency, where a study observed that 55% of patients with decreased DPD activity developed severe grade 4 neutropenia compared to 13% of patients who are DPD proficient[91]. About half of patients with DPD deficiency inherit germline genetic mutations from the currently screened risk variants of the *DPYD* gene, which encodes the DPD enzyme, leading to reduced DPD activity[4,91].

Table 1.5 shows four *DPYD* variants (c.1905+1G>A/DPYD*2A, c.2846A>T, c.1679T>G and c.1129-5923C>G/HapB3), which have been classified by the Clinical Pharmacogenomics Implementation Consortium (CPIC®) as having sufficient evidence to be used in clinical practice to screen for DPD deficiency to guide subsequent 5-FU chemotherapy dose modification[4]. Apart from c.1129-5923C>G, the rest of the variants have a frequency below 1% in European populations[91], and all four are rarer still or absent in Asian and African

populations. Three of the CPIC variants are missense mutations that alter the enzyme activity, while c.1129-5923C>G is a deep intronic splice site mutation that generates a non-functional transcript[71,92]. Patients who are heterozygous, compound heterozygous and homozygous carriers of these variants are at increased risk of ≥grade 3 toxicities due to reduced DPD activity[93]. The level of reduction in DPD activity differs in carriers of these CPIC-endorsed *DPYD* variants[71]. Variants c.1679T>G and c.1905+1G>A were observed to have the strongest and most deleterious impact on DPD activity, with a 68% and 50% reduction of activity in heterozygous carriers reported respectively[92]. In comparison, c.1129-5923C>G and c.2846A>T heterozygous carriers were reported to demonstrate moderate deleterious effects on DPD activity (35% and 31% reduction respectively)[92]. Patients who are heterozygote carriers for *DPYD* risk variants should receive a reduced dose in their initial 5-FU treatment due to incomplete DPD activity, while those who are homozygote carriers should avoid treatment with 5-FU[71].

Table 1.5 Four *DPYD* variants included in the screening test and listed in the CPIC® guidelines

***DPYD* variants**	**Effects from mutation**	**Estimated frequency in Europeans**	**Estimated effect on DPD activity**	**Sensitivity for ≥grade 3 toxicity**	**Dosing recommendations in heterozygous carriers**
c.1905+1G>A/*2A	Splice site mutation leading to skipping of exon 14 (encoding pyrimidine binding site)	0.79%	50%	5.2% (95% CI 3.0-8.9)	50%
c.2846A>T	Affects cofactor binding or electron transport	0.37%	31%	5.4% (95% CI 1.7-16.1)	50/75%
c.1679T>G	Destabilisation of protein structure	0.06%	68%	0.3% (95% CI 0.0-0.6)	50%
c.1129-5923C>G/HapB3	Abnormal pre-mRNA splicing	2.37%	35%	6.4% (95% CI 4.2-8.6)	75%

The *DPYD* variants have different effects on DPD and are present at different frequencies among the European population. The clinical applicability of screening for the *DPYD* risk variants is demonstrated via its effect on DPD activity and predictive ability of ≥grade 3 5-FU toxicity. Adapted from Meulendijks et al, *Cancer Treatment Reviews*, 2016[4,71]

A meta-analysis of eight studies involving 7365 patients found that variant c.1679T>G had the highest relative risk for 5-FU toxicity (RR 4.4, 95% confidence interval (CI) 2.1-9.3, p<0.0001)[71,94]. One study found that around only half of heterozygote *DPYD* variant carriers known to be associated with 5-FU toxicity are at risk of ≥grade 3 toxicity[95]. These findings indicate a probable compensatory and regulatory reaction towards the harmful effects of *DPYD* variants associated with 5-FU toxicity and divergence of DPD activity post-transcription[4,95]. As a result, the current *DPYD* screening test used in clinical practice is limited by its low sensitivity. Whilst it does detect patients at high risk of toxicity, it cannot identify all patients at risk of toxicity[71]. Patients can still develop severe toxicity due to other factors, including inheriting different rare deleterious *DPYD* variants, non-*DPYD* variants or other environmental factors[71].

To overcome the caveats related to the current *DPYD* test, there has been ongoing research on alternative mechanisms that could affect DPD activity. Several studies have explored the use of other DPD phenotypic and genotypic tests, including investigation on the effects of miR-27 microRNAs on DPD deficiency, that could help improve the predictive accuracy of the screening test[4].

1.7.2.2 Genetic factors predisposing to capecitabine toxicity

The genetic variations affecting the outcomes of patients exposed to IV 5-FU and oral capecitabine are anticipated to be similar, given that they share a similar metabolism pathway[89]. Alongside the *DPYD* variants mentioned above, genetic variants encoding for the activating enzymes of capecitabine to 5-FU are also potentially candidate genetic factors predictive of toxicity[89]. **Table 1.2** shows the genetic variants mapping to genes that encode the key 5-FU metabolising enzymes TS, MTHFR, DPD, and capecitabine activating enzymes

(CDA and CES) associated with capecitabine-induced toxicities that have been reported to date. Individual studies performed by Rosmarin et al. and Deenen et al. have found that variant rs2612091 mapping to *TYMS/ENOSF1* genes encoding for TS enzyme was associated with an increased risk of overall ≥grade 3 toxicities[12,96]. Two studies have also reported rs2612091 being linked with an increased risk of hand-foot-syndrome in colorectal cancer patients receiving capecitabine[12,97]. *DPYD* risk variant c.2846A>T, which is currently screened for before 5-FU treatment, was found to be predictive of capecitabine-induced diarrhoea toxicity[98,99]. Variant 79A>C mapping to the *CDA* gene has been reported by Garcia-Gonzalez et al. to be linked with increased risk of overall ≥grade 3 toxicity (OR= 1.84, P= 0.029) and associated with more ≥grade 3 haematological toxicities by Deenen et al[96,97]. Variants mapping to *ABCB1, ADCY2* and *OX40L*, which are genes involved in other pathways, have also been associated with capecitabine toxicity[97,100].

Risk variants can also be combined to improve their predictive ability. Rosmarin et al. observed an increased risk of ≥grade 3 capecitabine toxicity in carriers of *TYMS* variants 5'VNTR2R/3R and 3'UTR 6bp ins-del (combined OR= 1.31, P= $9.4x10^{-6}$) and *DPYD* variants (2846T>A and *2A) combined OR= 5.51, P= 0.0013)[101]. Many genetic variants associated with toxicity show inconsistencies between studies, which explains why few have been translated into clinical practice. Failure to validate variants is usually due to small discovery or validation cohorts and heterogeneity in sample populations and drug regimens. Although *MTHFR* 677C>T variant has been reported by Sharma et al. to be associated with ≥grade 2 HFS, other studies by Largillier and Rosmarin et al. did not find any significant association of this variant with capecitabine-induced toxicities[101–103]. The predictive ability of *CDA* gene variant 79A>C in identifying ≥grade 3 toxicity was not validated by a meta-analysis of 1177 patients by Rosmarin et al., of which no significant association with overall ≥grade 3 toxicity was observed (OR= 1.03, 95% CI= 0.86-1.24, P= 0.71)[101]. Variant *MTHFR* 1298A/A was reported by Sharma et al. to be

protective against grade 2 or 3 toxicity (OR= 0.06, 95% CI= 0.01-0.3, P<0.05). Still, it was also associated with more grade 3 to 4 haematological and GI toxicities in colorectal cancer patients from another study by Zarate et al[102,104]. Studies by Meulendijks, Zarate, and Deenen et al. have included patients treated with combination treatment of capecitabine and other chemotherapies, including oxaliplatin, docetaxel and irinotecan[96,104,105]. The lack of validated variants indicates the need for larger studies using a more standardised study population cohort and chemotherapy regimens.

Table 1.6 Previously reported genetic variants associated with capecitabine-induced toxicities

5-FU metabolising enzymes/Study	Study Population (sample size/cancer type)	Variant ID	Gene	Toxicity associations
Thymidylate synthase				
Garcia-Gonzalez [97]	239 CRC cancer	rs45445694 -2R-3R/3R-3R	*TYMS*	≥Gr 2 HFS (OR 3.63, 95% CI=1.18-11.22, P=0.025)
		rs2612091	*TYMS/ENOSF1*	>Gr 1 HFS (OR=2.28, 95% CI=1.10-4.76, P=0.027)
Rosmarin [101]	927 CRC cancer	5'VNTR2R/3R and 3'UTR 6bp ins-del	*TYMS*	≥Gr 3 overall toxicity (combined OR=1.31, P=9.4x10-6)
Meulendijks [105]	1605 (any cancers)	3RC/2RC (rs183205964), 2RG/2RC, 2RC/2RC	*TYMS*	Early global severe toxicity (OR=3.0, 95% CI=1.04–8.93, P=0.043)
Deenen [96]	34 upper GI cancers	1053C>T (C/C)	*TYMS*	Increased overall ≥gr 3 toxicity (P=0.004)
		rs2612091 (GG/AG)	*TYMS/ENOSF1*	Increased overall ≥gr 3 toxicity (P=0.027)
Rosmarin [12]	968 CRC cancer	rs2612091	*TYMS/ENOSF1*	≥Gr 3 overall toxicity (OR=1.59, 95% CI=1.39-1.79, P=5.28×10^{-6})
				≥Gr 3 HFS (OR=1.57, 95% CI=0.45-0.83, P=2.94×10^{-6})
Lin [106]	342 breast cancer	rs2606241	*TYMS*	Any grade HFS (OR=1.27, 95% CI=0.73–2.23, P=0.012)
		rs2853741 (C/T)	*TYMS*	Any grade HFS (OR=2.25, 95% CI=1.31–3.87, P=0.012)
Methylenetetrahydrofolate reductase				
Largillier [103]	105 breast cancer	677C>T	*MTHFR*	No significant association with ≥gr 3 global toxicity (P=0.54)
Sharma [102]	54 CRC cancer	1298A/A	*MTHFR*	Gr 2/3 toxicity (OR=0.06, 95% CI=0.01-0.3, P<0.05)
		677T/T	*MTHFR*	Gr 2/3 toxicity (OR= 0.1, 95% CI=0.01-1.0, P<0.05)
		677C/T and T/T	*MTHFR*	Gr 2-3 fatigue (OR=0.08, 95% CI=0.009-0.66, P<0.05)
		677C/T and T/T	*MTHFR*	Gr 2-3 HFS (OR=9.1, 95% CI=1.03-80.09, P=0.05)
Loganayagam [107]	244 (GI/unknown primary cancers)	1298C/C	*MTHFR*	Gr 2-3 HFS (OR=9.99, 95% CI=3.84–27.8, P=4.1×10^{-6})
Zarate [104]	60 CRC cancer	1298A/A	*MTHFR*	Increased gr 3–4 neutropenia (P=0.035), haematological (P=0.05) and GI toxicity (P=0.023)
Rosmarin [101]	927 CRC cancer	677C>T	*MTHFR*	Not predictive of ≥gr 3 overall toxicity (OR=1.10, 95% CI=0.89-1.35, P=0.39), ≥gr 3 diarrhoea (OR=1.09, 95% CI=0.79-1.49, P=0.60) or ≥gr 3 HFS (OR 1.07, 95% CI=0.85-1.35, P=0.58)

O'Donnell [100]	259 breast cancer	rs1801131	*MTHFR*	Early (cycle 1) and worst overall grade diarrhoea (OR=3.3, P = 0.02)
Lin [106]	342 breast cancer	rs3737964 (A/G)	*MTHFR*	Any grade HFS (OR=0.54, 95% CI=0.31–0.97, P=0.038]
		rs4846048	*MTHFR*	Any grade HFS (OR=0.54, 95% CI=0.30–0.98, P=0.042)
Dihydropyrimidine dehydrogenase				
Deenen [99]	568 CRC cancer	IVS14 +1G>A	*DPYD*	Severe toxicity (OR=12, 95% CI=1.4-109, P=0.004)
		2846A>T	*DPYD*	Severe diarrhoea (PPV 63%)
Joerger [98]	64 CRC and 76 upper GI cancer	c.85T>C	*DPYD*	Any grade HFS (P=0.033) any grade diarrhoea (P=0.023)
		c.2846A>T	*DPYD*	Any grade diarrhoea (P=0.028)
		c.1896T>C	*DPYD*	Any grade stomatitis (P=0.021)
Rosmarin [101]	927 CRC cancer	2846T>A and *2A	*DPYD*	≥Gr 3 overall toxicity (combined OR=5.51, P=0.0013)
O'Donnell [100]	259 breast cancer	DPYD 5* (rs1801159)	*DPYD*	Worst overall grade diarrhoea (OR 4.9, P = 0.0005)
Capecitabine-activating enzymes				
Caronia [108]	130 breast/CRC cancer	451C>T (rs532545)	*CDA*	Gr 3 HFS (OR= 2.02, 95% CI=1.02–3.99, P=0.039)
		943insC (rs3215400)	*CDA*	Gr 3 HFS (OR=0.51, 95% CI=0.27–0.95, P=0.028)
Loganayagam [107]	244 (GI/unknown primary cancers)	451C>T	*CDA*	Gr 2-4 diarrhoea (OR=2.3, 95% CI=1.3–4.2, P=0.008)
		92A>G	*CDA*	Gr 2–4 diarrhoea (OR=2.3, 95% CI=1.3–4.2, P=0.006)
Garcia Gonzalez [97]	239 CRC cancer	79A>C (rs2072671 A/A)	*CDA*	≥Gr 3 overall toxicity (OR=1.84, 95% CI=1.06-3.18, P=0.029)
Deenen [96]	34 upper GI cancer	79A>C	*CDA*	Increased grade≥3 haematological toxicity (P=0.038)
Martin [109]	192 breast cancer	823C>G (rs11075646)	*CES2*	≥Gr 3 HFS (OR=4.49, 95% CI=1.43-14.14, P=0.01)
Variants mapped to other pathways				
Garcia Gonzalez [97]	239 CRC cancer	Haplotype *1 (rs1128503 C, rs2032582 G, and rs1045642 C)	*ABCB1*	≥Gr 3 overall toxicity (OR= 4.06, 95% CI=1.97-8.38, P<0.001)
				≥Gr2 diarrhoea (OR=3.16, 95% CI=1.28-7.79, P=0.012)
O'Donnell [100]	259 breast cancer	rs4702484	*ADCY2*	Any grade diarrhoea (OR 3.0; P = 0.03)
		rs2205960	*OX40L*	Early (cycle 1) and worst overall grade HFS (OR=3.1, P=0.0007)

Summary of the genetic markers mapped to genes encoding for enzymes in the capecitabine metabolism pathway that have been reported to be associated with toxicity. Abbreviations: 5-FU- fluorouracil, CRC- colorectal, CI- confidence interval, GI- gastrointestinal, Gr- grade, HFS- hand-foot-syndrome, OR- odds ratio.

1.7.2.3 Association of MIR27A polymorphisms with fluoropyrimidine toxicity

MiR-27A and miR-27B are short single-stranded microRNAs involved in the post-transcriptional regulation of DPD activity[4]. These microRNAs help mediate the direct recruitment of RNA-induced silencing complex (RISC) proteins to *DPYD*, resulting in the degradation of *DPYD* mRNA and dampening of DPD expression[110]. A preclinical study found a significant negative correlation between DPD activity and miR-27A and miR-27b expression[110]. miR-27A also exhibits oncogenic properties and was found to be linked to an increased risk of metastases and chemotherapy resistance due to its indirect effects on drug efflux pump multidrug resistance protein 1[111–113]. The *MIR27A* gene encodes MiR-27A, and polymorphisms in this gene, such as rs895819 A>G, with a global minor allele frequency (MAF) of 35%, are associated with DPD activity[4,110]. rs895819 variant is located within the coding region for the hairpin structure of miR-27A, so it is hypothesized that this results in more efficient processing of miRNA precursors leading to more mature miRNA and higher miR-27A expression[114]. Functional studies in vivo supported this hypothesis; lymphoblastoid cell lines from carriers of the variant 'G' allele had significantly higher miR-27a expression (57% higher) than those from carriers of the wild-type 'A' allele[110]. In addition, carriers of rs895819 were found to have lower DPD activity in peripheral blood mononuclear cells than non-carriers[110]. These observations prompted further research on the impact of *MIR27A* polymorphisms on 5-FU chemotherapy toxicity.

Two studies to date have assessed whether the rs895819 variant is associated with 5-FU induced early-onset toxicity (CTCAE ≥grade 3 toxicity in cycles 1 and 2) as their study outcomes[20,111,115]. The first study by Amstutz et al. involved 514 patients receiving 5-FU-based chemotherapy and found a correlation between rs895819 and the risk of ≥grade 3 toxicities within the first two cycles[111]. However, this strongly depended on whether the

patients were *DPYD* low- or no-function variant allele carriers. A significant increase in risk was only found in the *DPYD* carrier cohort (OR= 7.6, 95% CI= 1.7-34.7, P= 0.0085)(47), and indeed a protective effect was found in patients carrying rs895819G who were non-*DPYD* risk variant carriers (OR= 0.6, 95% CI= 0.4-0.9, P= 0.012)[111]. 71% of patients who were carriers of both rs895819G and *DPYD* risk variants developed early ≥grade 3 toxicities[111]. Meulendijks et al. studied 1592 patients, the majority of whom (89.6%) were on capecitabine-based chemotherapy, and provided support for an interaction between rs895819G and *DPYD* risk variant genotype identifying that risk of toxicity in the first cycle of 5-FU based treatment was higher in rs895819G carriers who were also *DPYD* variant allele carriers (OR= 4.9, 95% CI= 1.2-19.7, P= 0.023)[115]. However, no association with risk was observed for rs895819G in patients not carriers of DPYD risk variants [115]. The same study by Meulendijks et al. also analysed the effects of a less common *MIR27A* variant rs11671784C>T (MAF 2%) and found no significant association with toxicity risk (OR= 1.1, 95% CI= 0.5-2.6, P= 0.777), although, given the rarity of this variant, the lack of association could be due to the study being underpowered[115].

A meta-analysis of the two studies above showed a significant association between rs895819 and fluoropyrimidine toxicity in carriers of *DPYD* risk variants (OR= 5.4, 95% CI= 1.8-15.7, P= 0.002)[115]. However, the number of patients with *MIR27A* variants who are also *DPYD* carriers from both studies is small (N=45 and 163 respectively)[111,115].

1.7.3 Genetic determinants of chemotherapy-induced cardiotoxicity

Host genetic factors could play a role in explaining the variation in susceptibility to cardiotoxicity from anti-cancer therapy seen among patients. Thus far, no validated genetic variants have been shown to be associated with 5-FU/Capecitabine-induced cardiovascular

toxicity. A meta-analysis of 41 studies involving 9,183 patients examined the association between 154 SNPs mapped to 88 genes linked to cardiac function[116]. Most of the studies included patients receiving treatment with anthracycline-based therapy. Six variants including rs4673 (*CYBA*), rs13058338 (*RAC2*), rs776746 (*CYP3A5*), rs45511401 (*ABCC1*), rs8187710 (*ABCC2*) and rs1136201 (*HER2*-Ile655Val) were found to associated with increased risk of cardiotoxicity[116]. rs1136201 had the highest effect size (OR=2.48, 95% CI=1.53–4.02, P=0.002) and had a low heterogeneity of I^2 =12%[116]. These genes mainly encode proteins involved in drug efflux and ROS defence and are thought to contribute to cardiotoxicity through the excessive accumulation of the drug and ROS formation[116].

1.7.4 Genetic factors influencing irAE development

1.7.4.1 Current evidence of genetic factors influencing irAE development

Several studies have analysed the association between genetic factors and irAE development. **Table 1.7** summarises the studies that have been conducted up to October 2022 and highlights variants that are significantly associated with irAE development.

Table 1.7 Genetic variants associated with irAE development from current studies

Author	Sample size	Cancer type	Type of ICI studied	Genetic analysis approach	Toxicity type	Chr	SNPs	Effect size (95% CI)	P value	Mapped gene/ nearest genes
Abdel-Wahab [117]	89 (44 cases, 45 controls)	Melanoma	Anti-CTLA4, anti-PD1/PDL1, combined ICI	All tagging SNPs	≥Grade 2 irAE	5	rs11743438	4.30 (2.3–8.0)	5.56×10^{-6}	*GABRP*
						5	rs11743735	4.50 (2.3–8.8)	8.34×10^{-6}	*GABRP*
						20	JHU_20.57183980	6.90 (2.7-17.6)	8.85×10^{-6}	*DSC2*
						2	rs56328422	4.20 (2.1-8.4)	4.14×10^{-5}	*BAZ2B*
						3	rs35807769	4.90 (2.2-11.1)	6.07×10^{-5}	Near genes *ZPLD1* and *MIR548AB*
						5	rs3026321	19.80 (2.6-152.7)	6.31×10^{-5}	*SEMA5A*
						5	rs1276216	6.40 (2.3-17.6)	8.18×10^{-5}	Near genes *LOC105374704* and *CDH6*
						2	rs919682	4.60 (2.1-10.1)	8.29×10^{-5}	*OSBPL6*
						2	JHU_2.179128562	4.60 (2.1-10.1)	8.29×10^{-5}	*AGPS*
						21	rs239731	11.30 (2.5-50.4)	8.56×10^{-5}	*LOC102724355*
						5	JHU_5.9269447	3.50 (1.9-6.5)	8.61×10^{-5}	*SEMA5A*
						2	rs359975	8.50 (2.4-29.9)	9.41×10^{-5}	Near genes *CFAP65* and *LOC100129175*
						3	rs11711517	0.16 (0.07-0.37)	2.39×10^{-6}	*LOC105377125*
						15	JHU_15.93602126	0.24 (0.13-0.46)	8.89×10^{-6}	*RGMA*
						15	JHU_15.93604000	0.24 (0.13-0.46)	8.89×10^{-6}	*RGMA*
						18	rs470753	0.16 (0.07-0.40)	1.65×10^{-5}	Near genes *LINC01924* and *CDH7*
						15	rs4778080	0.24 (0.12-0.47)	1.92×10^{-5}	*RGMA*
						12	rs2117997	0.22 (0.10-0.45)	2.44×10^{-5}	*ANKRD42*
						6	rs55733913	0.16 (0.06-0.42)	3.98×10^{-5}	*PACRG*
						12	rs7954686	0.16 (0.06-0.42)	3.98×10^{-5}	*FAR2*
						3	rs6440251	0.13 (0.04-0.40)	5.15×10^{-5}	*LOC105374140*
						14	JHU_14.101799748	0.11 (0.03-0.38)	5.20×10^{-5}	Near genes *LINC02285* and *LINC00524*
						18	rs4800887	0.25 (0.13-0.50)	5.74×10^{-5}	Near genes *CDH2* and *MIR302F*
						3	rs162263	0.24 (0.12-0.49)	5.88×10^{-5}	*ROBO1*
						9	rs10814859	0.24 (0.12-0.49)	5.88×10^{-5}	*GLIS3*
						3	kgp3960064	0.28 (0.15-0.53)	6.76×10^{-5}	*ROBO1*
						3	rs2062059	0.28 (0.15-0.53)	7.37×10^{-5}	Near genes *DCBLD2* and *MIR548G*

						8	rs12548560	0.28 (0.14-0.53)	8.95×10^{-5}	*PVT1*
						6	rs66502444	0.16 (0.06-0.43)	9.11×10^{-5}	*PACRG*
						8	rs6993547	0.16 (0.06-0.43)	9.11×10^{-5}	*PREX2*
Bins [118]	322 (96 cases, 226 controls)	NSCLC	Anti-PD1	Small-scale candidate study	Elevated liver transaminases	12	rs2301756*	2.31 (1.02-5.21)[a]	0.041	*PTPN11 (SHP2)*
					Rheumatological toxicity	12	rs2069705*	6.04 (1.53-23.86)[a]	0.019	*IFNG*
					Treatment-related adverse events	2	rs2227981*	0.45 (0.21-0.98)[a]	0.041	*PDCD1 804C>T (PD-1)*
Queirolo [119]	173 (10 cases, 163 control)	Melanoma	Anti-CTLA4	Small-scale candidate study	Endocrinopathies	2	rs4553808 (A/G or G/G genotypes)	0.09 (0.02-0.43)[a,b]	0.003	*CTLA-4 -1661A>G*
Cappelli [120]	753 (27 cases, 726 controls)	Mixed (largest cohort Melanoma)	Anti-PD1/CTLA4, Anti-PD1/ PDL1 +/- anti-CD73 or anti-LAG3	HLA specific approach	Inflammatory arthritis	6	HLA DRB1*04: 05	8.60 (1.70-43.40)	0.040	*HLA*
Magis [121]	163 (5 cases, 158 controls)	Melanoma	Anti-PD1	HLA specific approach	Acute type 1 diabetes	6	HLA DRB1*04:01, DRB1*04:05, DRB1*03:01	NA	NA	*HLA*
Refae [122]	94 (15 cases, 79 controls)	Advanced cancer (largest cohort NSCLC)	Anti-PD1/PDL1	Medium-scale candidate study	≥Grade 3 irAE	12	rs246079	0.09 (0.02-0.36)[c]	< 0.001	*UNG*
						9	rs10964859	6.08 (1.52-24.33)[b]	0.014	*IFNW1*
						9	rs4143815	8.11 (2.09-31.40)[b]	0.003	*PD-L1*
						19	rs12979860	3.72 (1.09-12.68)[b]	0.036	*IFNL4*
						2	rs3087243	3.19 (1.01-10.27)[b]	0.048	*CTLA4*
						2	rs11571302	4.06 (1.23-13.38)[b]	0.018	*CTLA4*
						2	rs7565213	3.73 (1.14-12.19)[b]	0.026	*CTLA4*
Iafolla [123]	101 (23 cases, 78 controls)	Mixed (largest cohort triple-negative breast cancer)	Anti-PD1	HLA specific approach	≥Grade 2 irAE	6	HLA-A	1.03 (0.23-4.62)[d]	0.970	*HLA*
						6	HLA-B	1.32 (0.24-7.26)[d]	0.750	*HLA*
						6	HLA-C	0.81 (0.24-2.78)[d]	0.740	*HLA*
						6	HLA-A, -B and -C	0.58 (0.21-1.62)[d]	0.300	*HLA*
Hasan Ali [124]		Advanced cancer	Anti-PD1	HLA specific approach	Pruritus	6	HLA-DRB1*11:01	4.53 ($X^2_{1,95} = 9.45$)	0.002	*HLA*

	102 (59 cases, 43 controls)				Colitis	6	HLA-DQB1*03:01	3.94 ($X^2_{1,95}$ = 5.67)	0.017	*HLA*
Luo [125]	729 (95 cases, 634 controls)	NSCLC	Anti-PD1 or combined anti-PD1 and CTLA4	All tagging SNPs	Thyroid disorders	6	rs9268543	NA	$7.5x10^{-7}$	*HLA*
Kirchhoff [126]	69 (30 cases, 39 controls	Melanoma	Anti-CTLA4	Whole exome sequencing and detection of ~1000 autoimmune variants	≥Grade 3 irAE	19	rs504963	2.57	0.006	3'UTR of *FUT2*
						6	rs1738074	2.21	0.025	*TAGAP*
Montaudié [127]	57 (33 cases, 24 controls)	Melanoma	Anti-PD1, Anti-CTLA4, combined anti-PD1 and anti-CTLA4	All tagging SNPs	Any grade irAE	2	rs782637	NA	4.10×10^{-8}	*PNPT1*
						2	rs78259	NA	3.40×10^{-7}	*PNPT1*
						2	rs7882572	NA	3.80×10^{-7}	*PNPT1*
						2	rs3762513	NA	6.60×10^{-7}	*CFAP36*
						11	rs7481951	NA	3.10×10^{-5}	*ANO5*
						2	rs4988956	NA	3.80×10^{-4}	*IL1RL1*
Yano [128]	11 cases	Melanoma, NSCLC, gastric	Anti-PD1, anti-CTLA4	HLA specific approach	Pituitary disorders	6	HLA-DR15	NA	1.40×10^{-3}	*HLA*
						6	HLA-DR1502	NA	0.002	*HLA*
						6	HLA-Cw12	NA	0.001	*HLA*
						6	HLA-B52	NA	0.003	*HLA*
Marschner [129]	167 (39 cases. 128 controls)	Melanoma, lung	Anti-PD1/PDL1	Small-scale candidate study	≥Grade 3 irAE	5	rs2910164 (C/C genotype)	6.78 (1.87–24.60)	0.004	*MIR146A*
Groha [82]	1751 (339 cases, 1412 controls)	Mixed (largest NSCLC)	Anti-PD1/PDL1, combined ICI, anti-CTLA4	All tagging SNPs	All grade irAE	8	rs16906115	HR=2.0 (1.6-2.5)	3.8×10^{-9}	*IL-7*

						1	rs75824728*	HR=1.9 (1.5-2.4)	8.4×10^{-9}	*IL22RA1*
						4	rs113861051*	HR=2.0 (1.6-2.6)	1.1×10^{-8}	Near genes *LINC02484* and *ARAP2*
Weidhaas [130]	161 (45 cases, 116 controls)	Mixed (largest cohort melanoma)	Anti-PD1/PDL1	Candidate study (50 common variants mapping to regions expected to disrupt miRNAs)	≥Grade 2 irAE	7	rs9374	AUC 0.82 (in a panel)	<0.001	*RAC1*
Udagawa [131]	622 (520 cases, 102 controls)	Mixed (largest cohort lung)	Anti-PD1	All tagging SNPs	All grade irAE	21	rs469490	5.15 (2.39-13.22)	2.97×10^{-7}	Near genes *APP* and *CYYR1-AS1*
					Thyroid irAE	15	Rs8023690	4.09 (2.21-8.15)	2.59×10^{-7}	Near genes *RGMA* and *LINC02207*

Summary statistics of existing studies reviewed up to October 2022 reporting on genetic factors identified as determinants of irAE. Effect sizes for all studies are reported in odds ratios and 95% CI unless indicated otherwise. Abbreviations: Chr- chromosome, CI- confidence interval, HR- hazard ratio, ICI- immune checkpoint inhibitor, irAE- immune-related adverse events, NA- data not available, NSCLC- non-small cell lung cancer, SNPs- single nucleotide polymorphism, [a]univariate analysis, [b]under recessive model, [c]under dominant model, [d]summary statistics based on HLA-1 loci heterozygosity relative to homozygosity *did not validate in independent sample sets

Most studies reported involved sample sizes of <200 patients and tested candidate genes linked with autoimmunity and the immune checkpoint pathways. The small sample sizes included limit the chance of identifying true associations. Bins et al. involved 322 non-small cell lung cancer (NSCLC) patients on nivolumab and under a recessive genetic model found associations between *PDCD1* 804C>T (rs2227981) and reduced risk of any grade toxicities (OR= 0.4, 95% CI= 0.2–1.0, P= 0.039), *PTPN11* 333-223A>G (rs2301756) with raised transaminases (OR= 2.4, 95% CI= 1.1–5.5, P= 0.036) and *IFNG* −1616T>C with rheumatological toxicity (OR= 6.0, 95% CI= 1.5–23.9, P= 0.019). However, these results could not be replicated in the validation cohort[118].

Given the evidence of an important role for *HLA* class 1 and 2 alleles in susceptibility to AD, they have also been investigated concerning irAE risk. As shown in **Table 1.7**, *HLA* variation explored thus far is associated with specific organ autoimmune toxicities rather than the overall risk of irAE. Significant associations were identified between *HLA-DRB1*11:01* and pruritus (OR=4.53, P=0.002) and between *HLA-DQB1*03:01* and colitis (OR= 3.94, P<0.017)[124]. *HLA-DRB1*11:01* and *HLA-DQB1*03:01* are linked with atopic dermatitis and colitis respectively[124,132]. *HLA* alleles are also associated with inflammatory arthritis (*HLA-DRB1*04:05*, OR= 8.6, 95% CI= 1.7-43.4, P= 0.04, n=26)[120] and T1DM (*HLA-DRB1*03* and *HLA-DRB1*04* haplotypes)[121]. Yano et al. found that *HLA-DR15* was significantly associated with pituitary irAE compared to healthy controls (81.8% vs 33.5%, P= 0.0014, n=11), where *HLA-DR15* has been linked with IL-17 mediated autoimmune diseases[128]. On the contrary, a study by Iafolla et al. that explored *HLA-1* haplotypes predicted from whole exome sequencing data for 101 patients treated with pembrolizumab did not find a significant association between any *HLA-I* loci or alleles with irAE development[123]. Mixed results have been observed for the development of ICI-induced T1DM. Association with *HLA DRB1*03/04*

group was identified in patients treated with anti-PD1; however, a case study of a patient treated with doublet ICI did not reveal any *HLA* risk allele associated with fulminant T1DM[121,133].

The first GWAS of irAEs involving 1751 patients predominantly on anti-PD/PDL1 identified a genome-wide association between a SNP mapping to *IL7,* rs16906115, and all grade irAE toxicities[134]. rs16906115 replicated in three other independent cohorts and is the first SNP linked with irAE development detected in a large study with validation cohorts[134]. Weidhaas et al. tested a panel of germline variants predicted to disrupt miRNA binding in 161 patients treated with anti-PD1/PD-L1[130]. The 50 variant panel, applied using four different classifiers, achieved an AUC of ~0.80 for predicting ≥grade 2 irAEs in their training and validation cohorts[130]. One of the markers in the classifier rs9374 was associated with a nine-fold increased risk of ≥grade 2 irAEs[130].

Groups of SNPs have also been tested for their ability to predict the risk of irAEs. Luo et al. generated PRS for hypothyroidism from GWAS of two thyroid-related phenotypes in UK BioBank; self-reported hypothyroidism and use of thyroid medication[125]. This PRS predicted thyroid irAE in NSCLC with a hazard ratio (HR) per standard deviation of 1.34, 95% CI 1.29-1.37 and area under the ROC curve (AUROC) of 0.6[125]. The authors also performed individual variant testing of 1502 hypothyroidism correlated SNPs identified in UK Biobank that were associated with thyroid irAE with nominal significance[125]. One SNP rs9268543, was associated with thyroid irAEs ($P= 7.5x10^{-7}$) after multiple test corrections, is located within the *HLA* locus and linked with various AD[125]. Exome sequencing studies have also been performed in small cohorts of patients treated with ICI. Montaudié et al. performed exome sequencing of 57 melanoma patients with the same number of patients available for

validation[127]. The authors concluded that coding germline variants have limited impact on irAE development; however, much larger cohorts are needed to establish this[127].

Most of the studies discussed above have been conducted in European populations. The first Japanese GWAS conducted by Udagawa et al. involving 622 patients identified 27 SNPs associated with any grade nivolumab induced irAE with a $P<1x10^{-4}$ [131]. 83.6% of patients in the study experienced an irAE, but only 86 and 16 controls were included in the discovery and replication phases respectively[131]. The top SNP, rs469490, identified from their combined discovery and replication analysis reached $P=2.97x10^{-7}$ (OR= 5.15, 95% CI= 2.39-13.22) and lies upstream to *APP,* which is downregulated in Crohn's disease[135]. A subgroup analysis identified rs8023690 as a potential predictive marker for hypothyroidism (OR= 4.09, 95% CI= 2.21-8.15, $P= 2.58x10^{-7}$)[131]. It would be helpful to test the performance of these SNPs in other ethnic populations, given the known variation of some AD correlated SNPs in different ethnicities.

1.8 Association between treatment toxicity and response

The relationship between chemotherapy or immunotherapy toxicity and response has been previously analysed. Abola et al. evaluated the toxicity and clinical data from 99 randomised phase 3 oncology trials to assess whether treatments with greater toxicity profiles than their respective controls had better responses[136]. The study found that cancer treatments that were relatively more toxic than their controls were associated with longer progression-free survival (PFS) but not OS[136]. When assessing specific toxicity types, the study found that gastrointestinal toxicity was significantly associated with higher PFS, which was not observed with haematological or cutaneous toxicities[136].

There has also been growing evidence that irAE occurrence correlates with increased treatment efficacy. A meta-analysis of 4324 patients found that irAE development was significantly associated with a reduced risk of death (HR= 0.49, 95% CI= 0.38-0.62, P<0.001), less risk of disease progression (HR= 0.51, 95% CI= 0.42-0.64, P<0.001) and increased odds of treatment response of 4.56 (95% CI= 3.72-5.59, P<0.001)[137]. A retrospective study involving NSCLC patients on nivolumab showed that patients with ≥2 irAE had markedly better survival outcomes than patients with one or no irAE (median PFS 8.5 vs 4.6 vs 2 months, median OS 26.8 vs 11.9 vs 4.0 months respectively, both P<0.0001)[138]. The same study also performed a 6 and 12-week landmark analysis which confirmed that early irAE occurrence at an early stage could be predictive of sustained clinical benefit[138]. Objective responses to pembrolizumab in melanoma patients were also linked to a higher occurrence of vitiligo[139]. Findings from these studies have identified a possible correlation between irAE and treatment efficacy; however, the underlying mechanism of this relationship is yet to be determined.

1.8.1 Genetic factors associated with treatment efficacy

Several studies have evaluated the predictive ability of genetic markers on fluoropyrimidine efficacy, particularly those affecting the drug's metabolism pathway. Variants of *TYMS* have been shown to explain differences in treatment responses between individuals. A study including 105 breast cancer patients on capecitabine found that *TYMS* 5' 3RG/3RG carriers had shorter PFS[103]. Joerger et al. also found that *TYMS* 3G promoter variants were associated with worse PFS in patients with advanced gastrointestinal cancers treated with combination capecitabine therapies[98]. Other studies have observed genetic variants associated with increased treatment efficacy. A study of 58 colorectal cancer patients found favourable responses with *TYMS* 5' 2R/2R[140]. Another study of 125 gastric cancer patients found

carriers of *TYMS* 3′ UTR del6 allele had longer median OS than carriers of *TYMS* 3′ UTR ins6/ins6 genotype (11.4 vs 6.8 months, P= 0.014)[141]. Apart from *TYMS* variants, Deenen et al. tested 8 *DPYD* SNPs in 568 patients on capecitabine, but none were associated with OS[99]. Studies have also assessed the relationship between polymorphisms of CDA and CES enzymes involved in capecitabine activation with treatment outcomes. A study of 136 patients found that patients carrying *CES2* 5′ UTR −823 G-allele had significantly better response rates (59 vs 32%, P= 0.015) and longer time to disease progression (8.7 vs 5.3 months, P= 0.014) than non-carriers. *CDA* variants rs602950 and rs2072671 have significant associations with PFS (HR per allele 1.44 and 1.77 respectively) in breast cancer patients treated with capecitabine alone[109].

Recent studies have demonstrated that autoimmune genetic risk loci could influence response to ICI. A study by Chat et al. involving 436 melanoma patients found that rs17388568 was most significantly associated with anti-PD1 response and carried a 74% lower chance of treatment resistance despite multiple testing adjustments (OR= 0.26; 95% CI= 0.12-0.53, P= 0.0002)[142]. This SNP maps to a locus containing the immune-related genes *IL-2, IL-21 and ADAD1*, implicated in autoimmune colitis, T1DM and allergy[142]. rs1893217 mapping to *PTPN2* was also strongly associated with response to anti-CTLA-4 therapy[142]. *PTPN2* has a negative regulatory role in cytokine signalling and has been linked with RA and T1DM[143,144]. Nomizo et al. found significantly better PFS in 50 NSCLC patients treated with nivolumab who were carriers of *PD-L1* variant rs2282055 G/G and G/T genotypes with a higher PFS of 2.6 months compared to 1.8 months in those carrying the T/T genotype (P= 0.016)[145]. The SNP has been associated with increased PD-L1 expression and several AD[145].

1.8.2 Are genetic variants associated with response also predictive of toxicity risk?

Given the potential linkage of treatment response and toxicity as discussed above, it would be interesting to find out whether variants associated with treatment efficacy are also associated with toxicity risk. Some genetic markers predictive of capecitabine toxicity have also been linked with treatment response. These include variants of *CDA* rs2072671 and *TYMS* 3'UTR6bp ins, which were associated with ≥grade 3 overall capecitabine-induced toxicity[101][97]. A retrospective study of 140 metastatic colorectal cancer patients on irinotecan and 5-FU found carriers of *ABCB1* 3435 were less likely to respond to treatment (OR= 0.32, 95% CI= 0.11-0.91) and were also at increased risk of early toxicity (OR= 3.79, 95% CI= 1.09-13.2)[146].

The presence of shared genetic susceptibility loci of ICI treatment outcomes has also been assessed. Refae et al. found no common SNPs between patients who responded to their anti-PD1/PD-L1 treatment and those who developed toxicity[122]. This study found that treatment response rates were significantly correlated with SNPs related to the tumour microenvironment, whereas SNPs associated with toxicity were predominantly target cell-related[122]. The top SNP linked with all-grade irAE from a large GWAS by Groha et al. was not associated with OS in their discovery cohort but was associated with improved prognosis in a cohort of melanoma patients from the Cancer Genome Atlas (TCGA) data[82].

1.8.3 Developing a predictive model of toxicity for use in clinical practice

Adverse short and long-term effects from chemotherapy and immunotherapy warrant further investigation into the development of a predictive model of toxicity to determine a patient's risk of developing ≥grade 3 adverse events. The ability to identify patients at high risk of a particular toxicity allows for stricter monitoring and earlier intervention to reduce the risk.

Indeed, the potential clinical benefit from their treatment also needs to be considered to not deprive patients of effective cancer treatment. Therefore, this balance between treatment efficacy and toxicity should be achieved by developing predictive models to estimate both treatment outcomes and help with risk stratification.

Conventional biomarkers used to estimate response to fluoropyrimidine based chemotherapy include tumour markers such as carcinoembryonic antigen (CEA) in colorectal cancer. However, more recently, the use of circulating tumour DNA (ctDNA) as a marker of response has been vastly studied. A study of stage 3 colorectal cancer patients evaluated the ctDNA status of patients who had post-operative ctDNA detected after adjuvant chemotherapy[147]. The study found that 23% of patients had ctDNA clearance after adjuvant treatment and did not relapse after 3 years, while all patients with transient or no ctDNA clearance relapsed[147]. Changes in ctDNA are also prognostic of treatment outcomes with a decrease in ctDNA correlating with increased PFS (HR= 3.29, P<0.0001) in 138 metastatic gastrointestinal patients[148]. This relationship has prompted large clinical trials to evaluate the ability of ctDNA to guide treatment options in colorectal cancer. Apart from germline variations, there are limited biomarkers of fluoropyrimidine chemotherapy toxicity. A study by Meulendijks et al. did find that high pre-treatment serum uracil concentration (>16ng ml^{-1}) was significantly associated with ≥grade 3 global 5-FU toxicity (OR=5.3, P=0.009) and gastrointestinal toxicity (OR= 33.7, P<0.0001) in patients treated with fluoropyrimidines[149]. However, this method is yet to be validated in an external cohort and involves labour intensive liquid-chromatography mass spectrometry for validation[149]. Considering the above evidence, prediction models incorporating host genetic variants, ctDNA, blood tumour markers, and tumour-related prognostic factors would therefore offer a personalised approach towards guiding management options for patients receiving fluoropyrimidine based treatments.

A range of predictive biomarkers of treatment response from ICI has been studied to date. The most widely used biomarker includes PD-L1 expression both within the tumour membrane and infiltrating immune cells, where a higher PD-L1 expression has been demonstrated to correlate with treatment response. However, relying on PD-L1 expression alone as a predictive marker of treatment response is inadequate due to the heterogeneity of PD-L1 expression within the tumour itself or between diagnostic specimens, differences in the detection assays used and inconsistent cut-off threshold for PD-L1 positivity for different ICI studied[150]. Other potential predictive biomarkers of response that are being evaluated include the level of tumour mutational burden, mismatch repair status, presence of tumour infiltrating lymphocytes, immune gene signatures, interferon-gamma related mRNA blood signatures and peripheral blood markers involving inflammatory cells such as neutrophils, lymphocytes and eosinophil count[150].

There has also been emerging research on biomarkers of irAE. Levels of cytokines and chemokines have been assessed as probable predictors of irAE. Lim et al. demonstrated an increased expression of pro-inflammatory cytokines such as IL-1a, IL-2, and IFNα2 before and during the early stages of treatment in melanoma patients who developed severe irAE[151]. Another study found high levels of IL-6 correlated with a higher frequency of severe irAE induced by anti-PD-1/PD-L1 treatment in NSCLC[152]. Pre-treatment IL-17 levels were also associated with severe colitis in melanoma patients receiving ipilimumab[153]. Although these studies were limited by their small sample sizes and retrospective nature, measurement of cytokine levels could be a practically helpful non-invasive biomarker of toxicity and more research is required to establish this.

Structural and functional changes of the adaptive and innate immune cells could also predict irAE. Oh et al. showed that ipilimumab induced early diversification of the T cell repertoire

with decline in T cell clonality within the first two weeks of treatment before the onset of an irAE[154]. This observation could explain the broad spectrum of irAE induced by anti-CTLA-4 therapy[154]. Early B cell changes were also correlated with a greater risk of irAE. Das et al. found that combined anti-CTLA-4 and anti-PD1 treatment reduced circulating B cells and increased $CD21^{lo}$ B cells and plasmablasts approximately three weeks before the onset of severe multi-organ irAE[155]. Deficiency in certain regulatory B cell phenotypes was found in patients who developed severe irAE[156]. The gut microbiome is also increasingly studied as a potential determinant of irAE, given the crucial role of the gut mucosal immune system in maintaining immune homeostasis[157]. Chaput et al. showed that baseline gut microbiota composition with *Faecalibacterium* and other *Firmicutes* had been associated with ipilimumab-induced colitis, whereas the presence of *Bacteroidetes* was protective against colitis[158].

Considering the evidence discussed above, generation of predictive models of ICI response and toxicity should incorporate both host genetic factors and other potential predictive biomarkers, including those from other 'omics' such as the microbiome, immunome, metabolome and the tumour microenvironment. The predictive ability of these biomarkers may also vary according to the tumour type and ICI therapy used; hence the prediction models used may need to be adapted to these factors. More research is still warranted to establish the optimal panel of biomarkers and assays that can be used in practice. Other factors that need to be considered include the cost-effectiveness, practicality of performing the test and patient's preference. Effective prediction models of treatment efficacy and toxicity can help clinicians decide on personalised treatment options to achieve maximal benefit from their cancer treatment while minimising the risk of serious toxicity.

As such, the main aims of my thesis are as follows:

1. To investigate novel genetic markers predictive of ≥grade 3 fluoropyrimidine chemotherapy-induced toxicities both during the first two treatment cycles and across all treatment cycles using genome-wide association studies (Chapter 3)
2. To assess the predictive ability of *MIR27A* variants in detecting patients at risk of early-onset capecitabine-induced toxicity (Chapter 3)
3. To determine the functional implications of a candidate gene associated with fluoropyrimidine induced cardiovascular toxicity (Chapter 4)
4. To detect potential genetic determinants of fluoropyrimidine related cardiac toxicity (Chapter 4)
5. To identify germline genetic markers of immune-checkpoint inhibitor-induced toxicity using genome-wide association analyses (Chapter 5)

CHAPTER 2: MATERIALS AND METHODS

2.1 Patient samples

2.1.1 QUASAR 2

The QUASAR2 clinical trial (International Standard Randomised Controlled Trial (ISRCTN) number: ISRCTN45133151) is an international, multi-centre randomised phase 3 trial that aimed to assess the benefits of adding bevacizumab to adjuvant capecitabine chemotherapy in colorectal cancer[159]. Patients were randomised to two arms: Arm A: Eight 3-weekly cycles of oral capecitabine (1250 mg/m^2 twice daily for 2 weeks and 1 week off) or Arm B: Capecitabine and 16 cycles of bevacizumab (7.5 mg/kg on the first day of each cycle) post-operatively. The National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) grading severity scale was used for reporting adverse events (AE) and is described in **Appendix 2**[20].

Whole genome tagging single nucleotide polymorphism (SNP) arrays, CTCAE graded toxicity data, and clinical data were available for analysis from 930 European stage 2 or 3 colorectal cancer participants of the QUASAR2 trial. A candidate gene-based analysis investigating maximum grade toxicity across all cycles has previously been published[12]. Full details of the genotyping arrays and imputation strategy are provided in the paper[12]. One of three Illumina tagging SNP arrays (Hap370, Hap610, Omni2.5) were used for genotyping and after quality control (QC), the data was imputed using the 1000 genomes August 2012 release (all ethnicities) and UK10K reference panel obtained from the European Bioinformatics Institute (EBI), dataset ID: EGAD00001000776. Access was granted via the electronic data access mechanism (EDAM) and data access agreements were signed with the Wellcome Sanger Institute.

~87.5 million (87,550,868) variants were available for analysis. Only variants with an imputation metric (INFO score)≥0.8 and a minor allele frequency (MAF) > 1% were considered

for analysis. After applying this QC filter, ~8.3 million variants were left for analysis. The *DPYD* variants listed by the Clinical Pharmacogenetics Implementation Consortium (CPIC) that were not included or could not be imputed from the tagging array data were genotyped using alternative methods[71]. Variants c.2846A>T and c.1679T>G were genotyped on the Illumina Exome Array, and c.1905+1G>A was genotyped using a custom KASP (LGC Genomics) targeted assay as described in Palles et al[160].

2.1.2 COIN

The COIN trial (clinicaltrials.gov ID: NCT00182715) was a randomised phase 3 trial assessing the effect of adding cetuximab to first line combination chemotherapy in advanced colorectal cancer patients[161]. The trial also investigated whether intermittent chemotherapy treatment is non-inferior to continuous chemotherapy. Patients were randomised to continuous oxaliplatin plus 5-FU/capecitabine (Arm A), oxaliplatin plus 5-FU/capecitabine with cetuximab (Arm B), or intermittent oxaliplatin plus 5-FU/capecitabine (Arm C). Oxaliplatin (130mg/m^2) and capecitabine (1000 mg/m^2 twice a day) were given 3-weekly. In Arm B, patients received a lower dose of capecitabine at 850mg/m^2 due to high toxicity rate. Oxaliplatin (85mg/m^2) plus 5-FU (bolus 400mg/m2 then infusional 2400mg/m^2 over 46 hours) and folinic acid were given two-weekly. Patients in Arm B received an initial dose of 400 mg/m^2 cetuximab and then 250 mg/m^2 once a week.

Genotyping and toxicity data were available for 1735 patients, with 998 patients treated with capecitabine-based chemotherapy. GWAS meta-analyses of COIN and QUASAR2 datasets to identify variants associated with cycle 1 or cycle 1 and 2 toxicity variables were conducted using METAv1.7. The same QC filters were applied to both datasets: MAF>1%, INFO score > 0.8, Hardy-Weinberg Equilibrium (HWE) > $1x10^{-6}$, and missingness <3%.

2.1.3 SCOT

SCOT is a randomised phase 3 clinical trial assessing the non-inferiority of 3 months of oxaliplatin with infusional 5-FU and folinic acid (FOLFOX) or capecitabine (XELOX) with 6 months of the same regimes in high risk stage 2 or 3 colorectal cancer[162]. The FOLFOX regime is given at a two-weekly combination of oxaliplatin (85mg/m^2) with 5-FU (bolus 400mg/m2 then infusional 2400mg/m^2 over 46 hours) and folinic acid, and the XELOX regime is given at three-weekly oxaliplatin (130mg/m^2) and capecitabine (1000 mg/m^2 twice a day for 14 days). Genotyping and chemotherapy toxicity data for 2939 patients were available for analysis.

2.1.4 Dutch cohort

Early-onset toxicity data (first two cycles) and *MIR27A* genotyping data were available for 1592 participants treated with fluoropyrimidine-based chemotherapy via collaborator Didier Meulendijks[115]. Data from this Dutch cohort excluded patients with the *DPYD* *2A variant as they were treated with a reduced dose of 5-FU in a previous prospective, multicenter study by the same authors (clinicaltrials.gov identifier NCT00838370) [163].

2.1.5 ICI GENETICS

ICI GENETICS is a national multi-centre case-control study aiming to identify genetic determinants of immune-related adverse events (irAE). Patients of any cancer diagnosis treated with ICI therapy from whom toxicity data is collected at each cycle are eligible for the study. The study's Chief investigator is Claire Palles, and the sponsor is University of

Birmingham (UoB). As of April 2023, the study was still open for recruitment. The study had appropriate ethical approvals from the Research Ethical Committee (REC) and Health Research Authority (HRA) in the UK with REC Reference 18/NE/0256 and Integrated Research Applications System (IRAS) Reference 237779. The study was also adopted by the National Institute for Health Research Clinical Research Network (NIHR CRN) portfolio of studies (Central Portfolio Management Services (CPMS) ID 39433).

Patients naïve to ICI are recruited prospectively before the administration of ICI or during their first two cycles of ICI treatment. Retrospective recruitment is possible for patients who have experienced a CTCAE ≥grade 3 irAE during their first line of ICI therapy. Patients' relatives diagnosed with autoimmune diseases can also be recruited into the study. **Table 2.1** shows the inclusion and exclusion criteria, and **Figure 2.1** shows the main recruitment process for the study.

Table 2.1 Eligibility criteria for enrolment to the ICI GENETICS study

Inclusion Criteria	**Exclusion Criteria**
1. Being treated for the first time with ICI(s) for any malignancy 2. Able and willing to provide informed consent and a sample of blood (or less preferably saliva) 3. Aged 16 years or over 4. Family members of participants with autoimmune disease are also eligible to participate, provided they are: • Able and willing to provide informed consent and a sample of blood (or less preferably saliva) • Aged 16 years or over	1. Previous exposure to ICI during the time of recruitment 2. Unable or unwilling to provide informed consent and a sample of blood (or saliva) 3. Aged less than 16 years old

List of inclusion and exclusion criteria for recruitment into the study for both prospective and retrospective cohorts. Abbreviations: ICI- immune checkpoint inhibitor

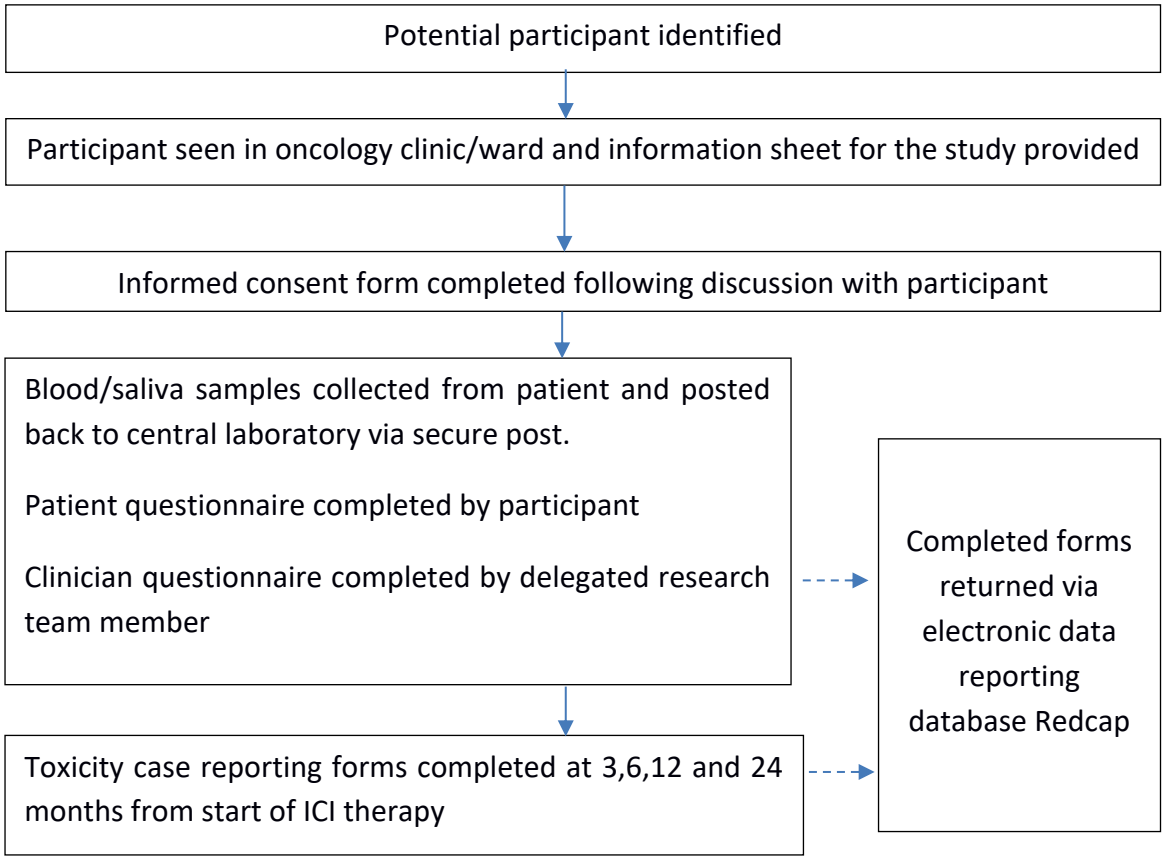


Figure 2.1 Flow chart demonstrating the recruitment process for ICI GENETICS
Study schema outlining the key steps involved when recruiting a participant into the study. Abbreviations: ICI-immune checkpoint inhibitor

2.1.5.1 Participating Sites

The West Midlands CRN promoted the ICI GENETICS study to seek potential recruiting sites from research teams based in England and Scotland NHS trusts. Expressions of interest were obtained from research teams keen to participate in the study. Following the sites' confirmation of Capacity and Capability, the green light was given to commence recruitment for the study. Site initial visit training was conducted for each study site via teleconferencing, and sample collection packs were supplied and posted to each site. Confirmation of research activities from each site was performed monthly via the NIHR CPMS portfolio, with annual reviews conducted from both the NIHR team and UoB research governance team as the study

sponsor. Delegated research team members involved in data collection were registered for access to the study project in the REDCap database and allocated to individual Data Access Groups for each recruiting site.

2.1.5.2 Sample Collection

Blood study packs containing two 3ml EDTA blood vacutainers, mailing plastic containers, blood collection equipment, and freepost padded envelopes labelled with UN3373 Biological Substance Category B labels were supplied and posted to the recruiting teams. Saliva kits containing OG-500 Oragene DNA Mailable Tube Format (DNA Genotek) for saliva sampling were also provided to the sites. Each sample collection pack was labelled with the participant's unique study ID. Participants were asked to provide approximately 6mls of blood or a saliva sample with their consent. The blood or saliva tubes were then placed in the mailing containers and posted back to the central research team using the Freepost envelopes provided.

2.1.5.3 Data Collection

Blood (or saliva) samples, consent forms, patient and clinician questionnaires were returned during the patient's initial visit. Clinical and toxicity information were collected using the REDCap (10.9.0) electronic database. Baseline clinical information requested in the questionnaires includes details of the patient's: baseline demographics, cancer diagnosis, cancer treatment, past medical history and drug history, personal and family history of autoimmune disease. Any documented toxicity resulting from the patient's ICI treatment was reported at study-specific observation periods at 3, 6, 12, and 24 months from the start date

of their ICI treatment. Information regarding the irAE, including the type of irAE, date of occurrence and recovery, maximum toxicity grade, treatment given for the AE and subsequent ICI management was recorded. Radiological response to treatment was also documented to assess the association between genetic factors and ICI treatment efficacy and to investigate the association between toxicity and survival.

2.1.5.4 Sample storage and DNA extraction

Sample and case report forms receipt was logged on a sample tracking database in a REDCap project set up by my lab group. Blood samples were aliquoted into 1.5ml micro-centrifuge screw-capped tubes and stored in separate -20 and -80°C freezers. Following the COVID-19 pandemic, the health and safety team at UoB required that we perform a heat incubation step at 60°C for 1 hour on all samples received before aliquoting and storing samples in the freezers. This additional step was done to inactivate any potentially active viral particles.

2.1.6 Discovery phase samples from IO-BASE consortium members

The IO-BASE (Immuno-Oncology agents- Investigating Biology to Address Serious adverse Events) consortium brought together researchers collecting samples, toxicity and survival data from patients treated with ICIs in the UK. Recruitment to ICI GENETICS had to be paused due to the COVID-19 pandemic in March 2020. As a result, we collaborated with members of IO-BASE who had collected toxicity data and blood samples from patients treated with ICI to increase the sample size in our discovery phase analysis to search for common variants associated with irAEs. ICI toxicity and clinical data were collated from our collaborators using a data collection template I designed. The data variables collected include age, sex, presence of ≥grade 3 irAE, maximum grade of irAE, cancer type, type of ICI agent, treatment for irAE,

follow-up period, date of irAE onset, any concurrent anti-cancer treatment, history of autoimmune disease and survival status. The study groups involved in the collaboration and the number of patient samples provided by each group are shown in **Table 2.2**.

Table 2.2 Studies of the collaboration that contributed to the initial GWAS of irAE

Study Name and Chief Investigators	Lead Site	No. of patients
PAIR (Sophie Papa, Andy Cope, Deborah Josephs)	King's College London	91
MOLGEN/HYST (Anna Olsson Brown)	Clatterbridge Cancer Centre, Liverpool	268
MEDALLION (Arthur Pratt, Ruth Plummer)	Northern Centre for Cancer Care, Newcastle	16
PRIMM (Julia Newton-Bishop)	Leeds Cancer Centre, Leeds	46
Oxford Radcliffe Biobank (Ben Fairfax)	Oxford	229
ICI GENETICS (Claire Palles)	University of Birmingham, Birmingham	169

The initial number of patients with DNA samples genotyped for the GWAS is listed. All samples have been genotyped using the targeted Global Screening Array v3. Abbreviations: GWAS- genome-wide association study, irAE- immune-related adverse event

2.1.7 Validation cohort

103 ICI GENETICS study samples were selected to form a validation phase cohort. All patients experienced one or more ≥grade 3 irAE. DNAs were extracted as per Section 2.11 and sent for genotyping as outlined in Sections 2.12 and 2.13. The control group (N=2497) were selected from UK Biobank (UK Biobank project reference: 86977). All had an ICD-10 coded diagnosis of melanoma, lung or kidney cancer. Genotyping and imputation data was generated using UK Biobank array- Affymetrix and IMPUTE2 v2.3.2[164] respectively using 1000 genomes and UK10K merged reference panel (dataset ID: EGAD00001000776) and Haplotype reference consortium panels (dataset ID: EGAD00001002729). A common set of SNPs in ICI GENETICS, UK Biobank and HapMap 3 was selected for identity by descent and principal component analysis (Section 2.14).

In silico validation was also performed for the top SNPs identified from the discovery phase GWAS in a cohort of 1751 patients of different cancer diagnoses treated with ICIs and of European ancestry[82]. Most (90%) of the patients were treated with anti-PD1/PD-L1; the rest were treated with the combination of anti-CTLA4 and PD-1 therapy. Genotype data were obtained from tumour biopsies sequenced using the OncoPanel sequencing platform. GWAS conducted from this cohort was based on Cox proportional hazard regression model, which considers time to toxicity event and survival time for each patient.

2.2 Coding adverse events for analysis

2.2.1 Coding capecitabine/5-FU-induced toxicity

Per cycle CTCAE graded toxicity data from 930 QUASAR2 participants were previously analysed to identify the maximum toxicity grade experienced across all treatment cycles. To analyse early events, I generated a binary variable for each commonly observed type of capecitabine-induced toxicities occurring in cycle 1 or cycles 1 and 2. Cases were coded as "1" denoting the presence of a ≥grade 3 toxicity, and controls as "0" who had no ≥grade 3 toxicity. The toxicities analysed were diarrhoea, neutropenia, thrombocytopenia, hand-foot syndrome, mucositis or stomatitis, and vomiting. Neutropenia and/or thrombocytopenia events were grouped under the haematological toxicity variable, and mucositis and/or stomatitis events were grouped under the mucositis variable. I also generated a global toxicity variable, where "1" denotes the presence of a ≥grade 3 event of any type of toxicities listed above and "0" if none of the toxicities were graded higher than grade 2. Cases for 5-FU induced cardiotoxicity was defined as patients who developed ≥grade 1 cardiac event. SNPTEST V2.5 (OXSTATEGEN, University of Oxford) format sample files were generated based on these coded variables[164].

2.2.2 Coding ICI-induced immune-related adverse events

Core variables including age, sex, type of ICI regime, date of first ICI treatment, presence of concurrent therapy, cancer diagnosis, and details surrounding ICI toxicity (irAE type, date of onset, maximum CTCAE grade, treatment of irAE) were collected for all samples included in the discovery phase cohort. Binary coded variables were generated for commonly reported irAE using a case-control approach to identify serious toxicity cases (CTCAE ≥grade 3 = case samples). The "serious toxicity" variable was generated based on the presence of any ≥grade 3 irAE (involving skin, gastrointestinal (GI), hepatobiliary, endocrine, thyroid, neurological, renal, musculoskeletal, blood and ocular toxicities), ≥grade 2 pneumonitis and any grade myocarditis. **Table 2.3** below demonstrates the types of adverse events induced by ICI for each organ-specific irAE variable.

Table 2.3 ICI-induced events reported for each organ-specific irAE variable

Skin	**GI**	**Endocrine**	**Hepatitis**
Rash +/- blistering Pruritus +/-rash Skin sloughing (SJS) Vitiligo	Diarrhoea/colitis Gastritis/duodenitis Pancreatitis	Hypothyroidism Hyperthyroidism Hypophysitis Type 1 diabetes Hypoadrenalism	Hepatitis Deranged LFTs
Respiratory	**Neuro**	**Renal**	**Musculoskeletal**
Pneumonitis	Neuropathy GBS Myasthenia Gravis Encephalitis Meningitis	Nephritis	Arthritis Myositis Arthralgia Myalgia
Cardiac	**Blood**	**Eye**	
Myocarditis	Haemolytic anaemia Thrombocytopenia Pancytopenia	Uveitis	

The types of irAE manifestations related to each organ system are shown here. Abbreviations: GBS- Guillain-Barré Syndrome, GI- gastrointestinal, LFT- liver function test, SJS- Steven Johnson's Syndrome

Age, sex, single or double-agent ICI therapy and the presence of concurrent anti-cancer treatment were included as covariates in association testing models. The "OST" score is an alternative method used to grade irAE and was devised by the PAIR study group (Kings College London). It takes into account the number of organs affected by toxicity ("O") and the number of CTCAE score ≥3 toxicities ("S") to generate the total cumulative toxicity score ("T"). OST scores were calculated for each participant included in the association analysis. Plotting a histogram of OST scores showed that the variable was not normally distributed. Square root and log transformations failed to yield a normally distributed variable, so I generated a categorical variable based on the individual scores (scores 8-14: high toxicity cases, 4-7: moderate toxicity, 1-3: low toxicity and 0: no toxicity as controls).

2.3 Statistical analysis

2.3.1 File formatting

To perform the genome-wide association analyses using SNPTESTv2.5[164], GTOOL (v0.7.5)[165] was used to convert KASP genotyping files or PLINK outputs in ped and map format to gen and sample file formats.

2.3.2 Frequentist based genome-wide association study analysis

To identify candidate genetic markers associated with CTCAE ≥grade 3 toxicity induced by fluoropyrimidines and ICI treatment, per variant association testing for each toxicity variable was performed using SNPTESTv2.5 based on frequentist testing for association. The

frequentist approach uses the existing database to test the null hypothesis that there is no significant association between single SNPs in the array with the toxicity phenotype tested, against the alternative hypothesis that an association exists[166]. Each analysis was adjusted for co-variables age and sex. To handle genotype uncertainty in association analyses from imputed data, missing data likelihood statistical theories are applied[167]. Association testing was initially carried out using the 'method score' option where SNPTEST calculates the observed data likelihood (contribution of each possible genotype weighted by its imputation probability) by maximising the first and second derivates of the likelihood under the null hypothesis[167]. Later, as described in the results chapters, I switched to using the 'method expected' option (uses expected genotype counts or dosage) to test for an association for a binary phenotype using a logistic regression model[167]. When the 'method expected' option is specified, the expected genotype count or allele dosage is analysed, and the genotype probabilities are renormalised to sum to 1[167]. Association testing for categorical toxicity variables was performed using a multinomial logistic regression likelihood model in SNPTEST v2.5.2.

2.3.3 Analysis of genome-wide association study data using PLINK

The GWAS of capecitabine toxicity analysis for the COIN trial dataset was performed by PhD student Katie Watts, Cardiff University. Imputed SNPs with a genotype uncertainty score (average maximum posterior call) of <0.9 were excluded. The probabilities for the rest of the imputed variants were hard calls generated using PLINK prior to association testing under the additive, dominant and recessive genetic models.

2.3.4 Analysis of genome-wide association study using the cox proportional hazard ratio

The Cox proportional hazards ratio model was used to perform a GWAS of any ≥grade irAE as unlike the logistic regression model in SNPTEST, in takes into account the amount of time taken to development of a ≥grade 3 irAE. This model may be more conducive given that the kinetics and onset time of irAE can vary between different irAE types and ICI regimes. These time to toxicity analyses were conducted using the R package gwasurvivr v1.14.0[168]. Serious CTCAE ≥grade 3 events were coded as events. Time to a ≥grade 3 event or time to last follow up was analysed.

2.3.5 Meta-analyses

GWAS meta-analyses was performed from one or more cohorts to improve the sample size of the population studied. The described meta-analyses for chemotherapy and ICI toxicity were conducted in METAv1.7 using the inverse-variance method based on the fixed-effects model. Outputs were filtered according to: HWE> $1x10^{-6}$, MAF>1% in cases and controls, presence in more than one trial or trial arm (n studies >1), GWAS significance ($P \leq 5x10^{-8}$) or suggestive significance ($P < 5x10^{-6}$) and P heterogeneity >0.05.

2.3.6 Variant annotation

Understanding the effects of the variants on expression of their mapped or nearest genes will allow us to gain more knowledge on how they influence toxicity risk. To determine the functional impact of the toxicity variants identified from our GWAS, the outputs from the analyses were annotated using gene annotation software tool ANNOVARv1.0 to[169].

2.3.7 Conditional analyses and joint association tests

The outputs identified from the GWAS can be in linkage disequilibrium (LD) with each other, which refers to the non-random association of alleles at different loci in a population which can happen due to their close distances from one another within the genome and mutation or recombination events[170]. Therefore, joint and conditional association analysis was performed using GCTA-COJO v1.93.2[171] to identify top independent toxicity SNPs that are not in LD with one another. Variants reaching genome-wide significance were first tested for independence using joint testing. Once a set of independent GWAS significant markers had been identified, all other SNPs were conditioned on these SNPs to determine whether there was evidence of second signals at any of the loci**.**

2.3.8 Gene Set and Tissue Expression Analysis

To understand the mechanism of toxicity development by the identified toxicity-associated genetic markers, gene-set analysis was performed to assess if these markers involved certain biological pathways contributing to toxicity. The gene sets being analysed involved groups of genes that share similar functional characteristics. Firstly, FUMA GWAS v1.4.1 was used to annotate summary statistics generated from META[84]. SNPs were mapped and assigned to protein-coding genes from Ensembl build 85. Gene-based analysis was then performed using curated gene sets and gene ontology (GO) terms from Msigdbv7.0. 10678 gene sets (curated gene sets: 4761, GO terms: 5917) from MsigDB v6.2 were used[172]. MAGMA gene-property analysis was performed to test the positive association between highly expressed genes in a specific tissue and disease-genetic associations using the average expression of genes per tissue type as a gene covariate. Gene expression values are log2 transformed average RPKM (reads per kilobase per million mapped reads) per tissue type based on GTExv8 RNA-seq

data[173], after winsorisation at 50 to mitigate the effects of outliers. The analysis was conducted based on the result of gene analysis (gene-based P-value) and tested for one side (greater) conditioning on the average expression across all the tissue types analysed. Individual tissue expression analysis was performed for 53 specific tissue types based on GTEXv8.

2.3.9 Interrogating publicly available datasets with expression quantitative trait loci (eQTL) and chromatin interaction information

Expression quantitative trait loci (eQTL) contributes to the genetic variation of a gene expression phenotype[174]. Following the identification of the top toxicity-associated SNPs, eQTL mapping was performed using FUMA GWAS v1.5.1 to understand the implications of the top toxicity-associated SNPS on changes in gene expression in specific tissues[84]. The data sources used for eQTL mapping include the following:

1. GTEXv8, which contains 54 different tissue types across 30 general tissue types[173]. It includes >17000 samples from 948 donors. Significant eQTLs are defined as False Discovery Rate (FDR) ≤0.05. The gene FDR is pre-calculated, and every gene-tissue pair has a defined P value threshold for eQTLs based on permutation.

2. eQTL catalogue, which contains multiple datasets across different tissue types. Significant eQTLs are defined as using a P value of 0.00001.

3. eQTLGen, which contains samples from meta-analyses of cis or trans-eQTLs from 37 datasets involving blood-derived gene expression of 31,684 individuals[175]. For cis-eQTLs, SNP-gene pairs with a distance <1Mb from the centre of the gene and tested in two or more

cohorts were included. For trans-eQTLs, only significant eQTLs were included. FDR was estimated based on permutations.

The genetic associations identified from GWAS can be challenging to interpret, particularly due to presence of high number of correlated variants in non-coding genomic regions. To prioritise genetic markers that are likely to be functionally important, analysing chromatin interaction data alongside eQTL data can be helpful in achieving this. The 3D conformation of chromatin in the nucleus allows long distance regulatory interactions between genomic regions. The characterisation of chromatin loops which connect the regulatory elements to their targeted genes, will allow us to better understand the genetic associations detected from GWAS. Hence the outputs from the GWAS was used to identify and map the chromatin interactions using FUMA GWAS v1.5.1 based on built-in chromatin interaction data obtained from Hi-C of 21 tissue or cell types from GSE87112 compendium (loops filtered at FDR 0.05 for multiple tissues)[84][176]. Hi-C loops were also identified from pre-processed enhancer-promoter and promoter-promoter interactions from adult and fetal brain cortex samples[177], where only significant interactions after Bonferroni correction (Pbon<0.001) were included. Hi-C based data from PsychENCODE, and Enhancer-Promoter correlations from FANTOM5 were considered.

2.3.10 Phenome-wide association study (PheWAS)

To improve our understanding of how the top toxicity variants detected from the GWAS influence the development of toxicity, phenome-wide association analyses can help explore the associations of the individual variants with a set of curated phenotypes using publicly available datasets[178]. PheWAS of the top variants were performed using the IEU OpenGWAS project database (https://gwas.mrcieu.ac.uk).

2.3.11 Polygenic risk scores

Complex disease traits such as toxicity development are often associated with one or more genetic markers[86]. To assess the predictive abilities of a panel of top independent SNPs in identifying patients at risk of ≥grade 3 toxicities, polygenic risk scores (PRS) were generated based on top SNPs identified from the GWAS. PRS are calculated for each individual based on the weighted sum of the risk alleles they carry using the summary statistics from GWAS data and accounts for the individual's genetic susceptibility to developing the toxicity analysed[87].

Polygenic risk score (PRS) analyses were performed using PRSice-2(2.2.2) to explore the sensitivity, specificity, positive predictive value and negative predictive value of PRSs based on toxicity-associated SNPs in predicting those at risk of ≥grade 3 toxicity[179]. PRSs were also derived from SNPs identified from the literature (including the National Human Genome Research Institute (NHGRI) EBI GWAS catalogue[56](last accessed April 2021) as associated with autoimmune diseases or immune traits from the cancer immunity cycle. For a SNP to be included in the PRS, the SNP had to reach GWAS significance ($P \leq 5x10^{-8}$) for the tested trait, not in linkage disequilibrium (LD) with other included SNPs, and have the following summary statistics available for analysis: effect allele, non-effect allele, beta (or OR) and p-value. Covariates used in previous association analysis, e.g. age, sex, single/doublet ICI agent and presence of concurrent therapy were included in PRSs. Genetic markers included in PRS were weighted by their known effect size – either odds ratio (OR) or beta taken from our data or published literature to generate the base data. LD clumping was performed using a r^2 threshold of 0.1, and ambiguous SNPs were removed.

Using the base data, PRSs are calculated for the individuals in the target sample dataset and tested for association with the target trait. The PRSs predictive abilities were determined using

the area under the receiver operating characteristic curve (AUC-ROC) using 'pROC' R package(1.17.0.1)[180]**.** Density plots showing the PRS distribution were generated using the 'car' R package(3.1-1**)**[181]**.**

To identify a set of common variants that can help predict ≥grade 3 irAE, I used LDpred-2 as an alternative method for polygenic risk scoring based on a Bayesian approach[182]. The analysis was restricted to HapMap3 SNPs. The GWAS summary statistics data from the discovery phase was loaded, and the LD correlation matrix was computed from UK Biobank genotype data of European ancestry as reference. LD score regression was performed to obtain the SNP heritability score and tested in three PRS models.

2.3.12 Receiver operating curve analysis

To assess the predictability of a PRS model in identifying patients at risk of ≥grade 3 toxicity, area under the receiving operative curve (AUC-ROC) analysis can be conducted based on the PRS scores calculated for each individual. R package 'ROCit' v2.1.1 was used to explore the AUC, sensitivity, specificity, positive predictive value and negative predictive value of smaller SNP sets (1-2 markers, e.g. *MIR27A* SNPs in Chapter 3).

2.3.13 Transcriptomic wide association analysis (TWAS)

Identifying the causal genetic marker leading to toxicity development can be challenging based on GWAS data alone. Transcriptomic wide association analysis (TWAS) is therefore used to help identify gene-trait associations based on eQTL data to identify candidate genes of toxicity[183].

A gene-based association cross-tissue transcriptomic analysis was performed using SPrediXcan from the MetaXcan software with the discovery GWAS summary statistics of ≥grade 3 irAE. The multiadaptive shrinkage in R (MASHR) model database, containing 49 GTEX

tissues, was used as the transcriptomic prediction model. The GTEXv8 database provided the covariance matrices of SNPs within each gene model. A P-value threshold of $<2.5x10^{-6}$ (0.05/20000) genes was considered significant.

2.3.14 Power calculations

An in-house power calculation tool, kindly shared by Doug Altman was used to assess the effect sizes powered at 80% given the sample size and frequency of the test alleles.

2.4 KASP Genotyping

Both *MIR27A* SNPs rs895819 and rs11671784 were imputed in QUASAR2 with INFO scores of 0.96 and 0.62 and missing data rates of 7% and 5% respectively. Hence, I decided to directly genotype these SNPs using KASP (LGC Genomics) genotyping assays to improve the genotyping accuracy and minimise the data missingness.

Assays were designed to genotype rs895819 and rs11671784 using Kompetitive allele specific PCR (KASP). The SNP sequences were obtained from UCSC Browser (https://genome.ucsc.edu) and submitted to LGC Genomics for KASP SNP assay design[184]. DNA samples from QUASAR2 had previously been normalised to 20ng/µl and arrayed in 96-well PCR reaction plates with wells E12, F12, G12, and H12 left blank for controls. 1µl of each DNA was used for genotyping. Genotyping reactions were set up in FAST plates (Cat no. 4413255). One no template control (NTC) and positive controls corresponding to each observed genotype (homozygous for allele A (A/A), homozygous for allele B (B/B) or heterozygous for allele A and B (A/B) were included in each plate in wells E12-H12. The SNP assays designed by LGC were used in combination with 2X KASP Master Mix (KBS-1050-122, Low ROX 5000 v4.0TF) and the amounts used per

10µl reaction are shown in **Table 2.4**. After mixing 9µl of master mix with each DNA sample, the reaction plate was sealed with an optical lid and centrifuged. The genotyping reactions were then run under thermal cycling conditions according to the KASP thermal protocol: 94°C for 15 minutes, 94°C for 20 seconds, 61-55°C for 60 seconds (drop 0.6°C /cycle) for 10 cycles, 94°C for 20 seconds and 55°C for 60 seconds for 26 cycles.

Table 2.4 KASP genotyping mix components

Reagent	**Volume (µL per well)**
2X KASP Master Mix	5
KASP Assay mix	0.14
Water	3.86
DNA sample	1
Total reaction volume	10

After completing thermal cycling, the reaction plate was read using the Applied Biosystems QuantStudio 5 Real-Time PCR system. Genotype calling was performed using ThermoFisher Scientific Design and Analysis Software v1.5.2 with manual review and reassignment of genotypes if required. Repeat thermal cycling for 10 more cycles was performed if the genotypic clusters were not well-defined. For each SNP, genotyping was repeated for samples with undetermined results. 5% of the total number of samples were randomly selected to check for concordance. Genotype data generated from imputation and KASP genotyping were compared using the linkage format in HAPLOVIEW to check for concordance[185].

2.5 Polymerase chain reaction, PCR clean up, agarose gels and Sanger sequencing

Sanger sequencing was used to confirm KASP genotyping results of the two *MIR27A* variants. Samples were selected for each observed genotype (homozygous/heterozygous for risk allele). The DNA sequence surrounding each SNP was obtained from the UCSC browser. Primers were designed using Primer3Plus and synthesised by Merck Life Science (25nM scale, desalted purification). Polymerase chain reactions (PCR) for each sample were set up as shown in **Table 2.5**.

Table 2.5 PCR master mix (10µl reactions)

Reagent	Volume per sample (µl)
DNA sample	1
Amplitaq gold 360 (2X)	5
Primer F (100µM)	0.2
Primer R (100µM)	0.2
GC enhancer	1
Water	2.6

PCR was performed according to the Amplitaq gold protocol: 95°C for 10 minutes, followed by 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute (repeated 35 times) and 72°C for 7 minutes. The annealing temperature in step 3 was altered from 57°C to 60°C to optimise the PCR.

Agarose gel electrophoresis was performed to check the amplification quality of the DNA samples. 1X Tris-Borate-EDTA (TBE) buffer is mixed with 1% agarose powder (Sigma-Aldrich) to form the gel. The gel mixture is heated in the microwave at high power until it is clear and placed in the heat incubator. 1X SYBR Safe DNA gel stain (Invitrogen™) is added to the mixture, and once sufficiently cooled to room temperature, the mixture is poured into the gel tray, and an 8-well comb is placed at the end of the tray. Once the gel mixture has solidified,

the tray is placed in a water bath filled with 1x TBE buffer and the comb is removed. 1kb Plus DNA Ladder (Invitrogen™) was used as a marker to approximate quantification of the DNA samples. Orange 5X loading dye was mixed with each DNA sample and the ladder before loading the samples onto the gel. The gel was run at 120V for 30 minutes and the DNA fragments were checked under the UV light.

2µl of ExoSAP-IT™ *Express* PCR Product Cleanup Reagent (Applied Biosystems ™, Thermo Fisher Scientific) was then added to 5 µl of each PCR product and incubated at 37^0C for 4 minutes to hydrolyse unused primer and nucleotides and at 80^0C for 1 minute to inactivate the enzymes. Purified PCR products were diluted by adding 10µl water before the cleaned PCR reactions and primers (diluted to 3.2pmol) were sent for Sanger sequencing by Source BioScience Limited. The sequencing data generated was read using FinchTV (v1.5, Digital World Biology) and genotypes observed from sequencing were compared against that from KASP genotyping.

2.6 Cell culture experiments assessing the impact of 5-FU and *DGLUCY* knockdown on mouse cardiomyocytes

2.6.1 Preparing HL-1 cell media

Collaborators from the Institute of Cardiovascular Sciences, UoB, kindly provided mouse cardiomyocyte cell lines (HL-1). These cells are derived from the AT-1 mouse atrial cardiomyocyte tumour lineage. Culture media for these cells is supplemented Claycomb media without L-glutamine (Sigma-Aldrich, Cat no. 51800). Contents of the medium include Claycomb Medium, 10% Fetal bovine serum, 100U/ml:100ug/ml of Penicillin/Streptomycin,

0.1mM of Norepinephrine and 2mM of L-glutamine. The media is wrapped in aluminium foil and stored in the fridge as it is light-sensitive.

2.6.2 Cell culture

HL-1 cells were cultured in T25 flasks (Corning® cell culture flasks) coated with 0.02% Gelatin/Fibronectin. Cells were incubated at 37^0C in a cell culture CO_2 incubator. T25 flasks were coated with 0.02% Gelatin/Fibronectin for at least 1 hour before splitting or plating the cells. Media was removed from the flask containing the cultured cells. ~3ml of 1X Phosphate Buffered Saline (PBS) was used to wash the cells. 3ml of 0.05% trypsin was added to the flask and returned to the incubator for 3 to 5 minutes. Cells were visualised under a microscope to check for cell detachment. Once confirmed, the trypsin was neutralised by adding 5mls of media. The media/trypsin suspension was transferred to a 15ml Falcon tube and centrifuged at 700 x g for 5 minutes to pellet the cells. The supernatant was removed via aspiration and the cell pellet was resuspended in ~3ml of media to produce a single-cell solution for cell counting.

2.6.3 Cell counting and plating for metabolite extraction

10µl of cell suspension was collected in a 1.5ml Eppendorf tube. 10µl of trypan blue was added and resuspended with the cells, and this was then loaded onto the hemacytometer. The concentration of cells was measured by counting the number of cells within 2-3 large squares of the haemocytometer chamber under a microscope and using that information in the following formula:

Concentration = (No. of cells counted/No. of squares counted) x 0.5 (dilution factor) x 10^4 /ml

2.6.4 Metabolite harvest and extraction for gas chromatography-mass spectrometry (GC-MS)

In collaboration with Professor Daniel Tennant's group at The Institute of Metabolism and Systems Research, GC-MS was conducted by the group to detect D-glutamate, the metabolite product of D-glutamate cyclase. The reagents and conditions required for the extraction of L-glutamate and D-glutamate metabolites are:

a) 25mM pentanedioic acid-d_6 (D6-IS) in methanol (3.5mg/ml stock). D6-IS is further diluted 1:1000 in water and stored at 4°C.

b) Methanol (MeoH) and chlorofoam ($CHCL_3$) solutions stored separately at -20°C.

Cells in a 9-well plate were washed with PBS and placed on ice. 500µl of methanol was added to each well. A cell scraper was used to detach cells and the suspended cells were transferred to 2ml Eppendorf tubes. These tubes were centrifuged briefly at 1 x g for 10 seconds at 4°C to ensure no liquid remained on the Eppendorf caps. 200µl of D6-IS/water solution and 500µl of $CHCL_3$ are added to each tube and vortexed for 10-15 seconds. The tubes were then centrifuged at 14,800 x g for 10 minutes at 4°C resulting in two phases of solution in each tube. The upper phase of the solution from each tube was transferred to 2ml Eppendorf tubes. The solution in the tubes was then dried using a vacuum concentrator for 2 hours and the resultant solute was subsequently stored in a -20°C freezer in preparation for GC-MS. Cell numbers ranging from 50,000 to 500,000 cells were sent for analysis using GC-MS. Untreated cells and cells treated with 0.1µM 5-FU were analysed.

2.6.5 Treatment of cells with 5-fluorouracil

Helen Robbins kindly provided 5-FU from the Beggs group, Institute of Cancer Genomics, UoB. The stock concentration was 100mM. The dilutions required for treating the cells for viability of RNA analysis (0.1µM, 0.25µM, 1µM, 5µM, 10µM, 25µM, 100µM, 200µM and 400µM) were made up by diluting the stock drug with Claycomb media. Drug treatment experiments were conducted in 96-well plates coated with gelatin/fibronectin. 3000-5000 cells were seeded per well and drug concentrations were each tested in five replicate wells. Cells were treated once or twice (24 and 96 hours after the cells were first plated) before adding MTS assay and measuring absorbance results[186]. The Beggs group previously used this treatment regime when treating colorectal cancer cell lines with 5-FU. Untreated cells acted as the negative control.

2.6.6 Measuring cell viability

One of the aims of the cell culture experiment was to assess the cell viability of HL-1 mouse cardiomyocytes following 5-FU treatment. In this experiment, 20µl of the MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) assay colorimetric kit reagent was added to each well of a 12-well plate containing 70,000 cells that were treated with 5-FU for 24 hours. Reduction of the MTS tetrazolium compound by viable cells generates a formazan dye that is quantified by measuring the absorbance at a wavelength of 490nm. The absorbance results of each well were measured at 15 minutes, 1 hour and 4 hours after incubation with MTS using a 96-well plate reader with a shaker function.

.

2.6.7 Knockdown of *DGLUCY* in cells

2.6.7.1 Transfection reactions

To determine the knockdown effects of DGLUCY in HL1 cells, transfection experiments were set up using small-interfering RNA (siRNA). siRNA targeting *DGLUCY* were designed from Mouse 9030617O03Rik siRNA using ONTARGET*plus* SMARTpool technology by Dharmacon Horizon Discovery Bioscience Limited. ONTARGET*plus* non-targeting pool of siRNA were used as controls. HL-1 cells were plated in a 6-well plate (170,000 cells/well). ~24 hours after plating, when the cells were visualised to be ~60-80% confluent, I transfected the cells according to the RNAiMAX transfection procedure. In this procedure, 9µl of Lipofectamine RNAiMAX® reagent is diluted in 150µl of OPTI-MEM® Medium. 3µl and 6µl of siRNA ("Scramble/SCR" untargeted pool of siRNA or targeted siRNA pool) is also diluted in 150µl of OPTI-MEM® Medium. To test the optimal concentration of siRNA needed for effective transfection, both 30pmol (3µl) and 60pmol (6µl) were tested. The diluted siRNA was then diluted in Lipofectamine RNAiMAX® reagent at a 1:1 ratio and incubated at room temperature for 5 minutes. 250µl of the combined siRNA-lipid complex is then added to each well. The procedure was repeated for all six wells, with the top three wells transfected with untargeted 'SCR' siRNA and the bottom three wells transfected with siRNA targeting *DGLUCY.* In addition, 2.5µl of Green Fluorescent Protein (GFP) reporter was added to one set of wells testing for 30pmol of siRNA as a technical control to determine transfection efficiency. The transfected cells were then incubated for 24-72 hours at 37ºC.

2.6.7.2 Optimising transfection experiments

24 to 72 hours post-transfection, the cells were harvested from each well in preparation for RNA extraction (Section 2.8), conversion to complementary DNA (cDNA) (Section 2.9) and

gene expression analysis as described in Section 2.10 using Taqman Assays for *DGLUCY* (target gene), *RPL32* and *PGK1* (housekeeping genes). The optimal time required to achieve maximum knockdown of *DGLUCY* was determined by analysis of *DGLUCY* gene expression at 24, 48 and 72 hours.

2.6.7.3 Assessing cell viability of DGLUCY knockdown cells post 5-FU treatment

100,000 HL-1 cells were plated into 6-well plates. ~24 hours after plating, duplicate wells were transfected with 30pmol untargeted pool 'SCR' siRNA or 30pmol *DGLUCY* siRNA. ~24 hours post-transfection, cells from each well were scrapped and collected using a cell scrapper and centrifuged ready for seeding wells of 96-well plates, as shown in **Figure 2.2**.

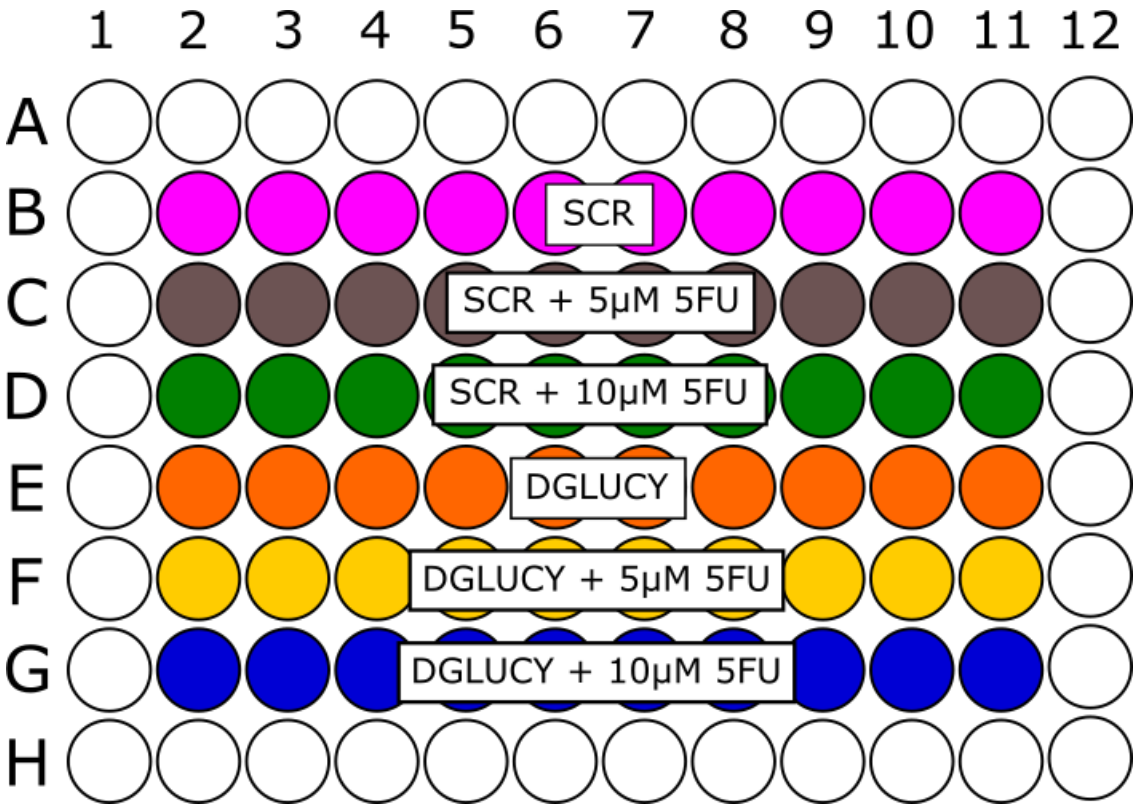


Figure 2.2 96-well plates containing siRNA transfected HL-1 cells
Each of the 6-well plate wells seeded 10 wells of two 96-well plates. Each well was seeded with 2000 cells. 48 hours post-transfection, 'SCR' and DGLUCY siRNA transfected cells in both 96 well plates were treated with 5μM or 10μM of 5-FU.

MTS assay was performed 24 hours after the first dose of 5-FU (or three days post-transfection) for samples in the first 96-well plate, which was subsequently discarded. For samples in the second plate, a second dose of 5-FU was administered 48 hours after the first dose of 5-FU was given using the same drug concentration. MTS assay was then performed 48 hours after the second dose of 5-FU treatment (or six days post-transfection).

2.6.7.4 Analysing gene expression of DGLUCY knockdown transfected cells post 5-FU treatment

Another part for functional analysis of *DGLUCY* was to determine the level of expression of *DGLUCY* in HL-1 cardiomyocytes following 5-FU treatment. Part of the experiment involved testing for *DGLUCY* gene expression of cells post-transfection with siRNA alongside the above cell viability experiment. Two wells of 'SCR' or DGLUCY siRNA transfected cells were plated in a 12-well plate to test for their gene expression at three and six days post-transfection. During this process, cells were harvested with buffer RLT and stored at -20ºC three and six days post-transfection to match the same period when cell viability was assessed via MTS assay. Real-time qPCR was performed as per Section 2.10.

A summary of the timeline of which both experiments in Section 2.6.7.3 and 2.6.7.4 was performed is shown in **Table 2.6**.

Table 2.6 Timeline in the steps taken to assess cell viability and gene expression of transfected HL-1 cells treated with 5-FU

Timeline	Steps of experiment
Day 1	Plating of HL-1 cells
Day 2	Cells transfected with 'SCR' and *DGLUCY* siRNA.
Day 3	Cell harvested, counted and plated in a 96-well plate in preparation for MTS assay and in parallel for qPCR to check for gene expression
Day 4	Cells treated with first dose of 5-FU
Day 5	MTS assay performed. Cells harvested to check for qPCR expression.
Day 6	Cells treated with second dose of 5-FU
Day 7	Observation
Day 8	MTS assay performed. Cells harvested to check for qPCR expression.

The days from initial transfection and 5-FU treatment to MTS assay and RT-qPCR of HL-1 cells are outlined. 'SCR' or scramble cells refer to an untargeted pool of siRNA. Abbreviations: 5-FU- fluorouracil, qPCR- quantitative polymerase chain reaction, siRNA- small interfering RNA.

2.7 Mitochondrial isolation for detection of D-glutamate

Previous literature described isolating mitochondria from rat cardiomyocytes to detect the D-glutamate metabolite[187]. Digitonin was used to selectively permeabilise the cytosolic membrane of cells while keeping the mitochondrial membranes intact[188].

2.7.1 Digitonin concentration optimisation

The digitonin concentration was first optimised to determine the concentration that can induce complete permeabilisation of cells without permeabilising the mitochondria membrane. HL-1 cells were plated in a 6-well plate to test five different digitonin concentrations for this procedure. A negative control untreated well was also prepared. A mitochondrial buffer was used to dilute the digitonin. The buffer was prepared on the day of use and it consisted of: 2mM Potassium phosphate monobasic, 120mM Potassium chloride,

3mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)), 1mM ethylene glycol-bis tetraacetic acid (EGTA) and 3g/l Bovine serum albumin (BSA). The pH of the buffer was adjusted to 7.2 using KCl and NaOH. The culture media was removed once the cells reached the desired confluency of ~70-90%. The cells are then washed with 500µl saline (0.9% NaCl) solution.

500µl of mitochondrial buffer with different concentrations of digitonin to be tested (control 0µg/ml, 25µg/ml, 50µg/ml, 75µg/ml, 100µg/ml and 300µg/ml) was added to individual wells. The plate of cells was then incubated for 2 minutes at 37ºC. The mitochondrial buffer and digitonin solution were removed from each well, which was then washed with 500µl of saline. Trypan blue was added to each well and the cells were checked under microscopy to assess if the cells were stained blue (cells are permeabilised). The lowest concentration of digitonin, which caused the cells to stain blue, was then used for subsequent digitonin cell fractionation process for mitochondrial isolation.

2.7.2 Digitonin cell fractionation to isolate mitochondrial from transfected and non-transfected HL-1 cells

2.7.2.1 Non-transfected cells

HL-1 cells were plated in six T25 flasks (~200K cells/flask). After a week of incubation, the cells were treated with digitonin for mitochondrial isolation. In this process, the mitochondrial buffer was prepared on the day of the procedure as described in Section 2.7.1 and mixed with the selected optimal digitonin concentration. Media was removed from the cells, which were then washed with saline. The combined mitochondrial buffer and digitonin solution were added to each flask and then incubated at 37ºC for 2 minutes. The flask was washed again with saline.

500µl of combined methanol and D6 glutaric acid (diluted 1:1000 in methanol) solution was added to each flask to isolate the mitochondrial fraction of the cells, which was harvested into a clean 2ml Eppendorf tube. The same process was repeated for all six flasks. The volume collected from two flasks was combined into one Eppendorf tube, resulting in three Eppendorf tubes representing a triplicate of cells treated with the same conditions. 1ml of chloroform was added to the three tubes containing the mitochondrial fraction. The tubes were then vortexed and centrifuged. The upper phase of the post-centrifuged solution was then dried in Eppendorf Concentrator for 2 hours as per protocol in Section 2.6.4 for GC-MS to detect L- and G-glutamate.

2.7.2.2 Transfected cells

A 6-well plate of ~100,000 HL-1 cells per well was prepared. After 24 hours, triplicate wells were transfected with 'SCR' siRNA or siRNA targeting *DGLUCY* as per Section 2.6.7.1. After 48 hours, cells from each well were transferred to six T75 flasks to expand the cell numbers. After a week of incubation, mitochondrial were isolated from the cells using the digitonin cell fractionation protocol described in the previous section. After removing the media and washing the cells with saline, ~1ml of combined mitochondrial buffer and digitonin solution is added to each flask and incubated for 2 minutes. After washing with saline, 1ml of combined methanol and D6 glutamic acid solution is added to each flask and scraped to collect the mitochondrial fraction into clean Eppendorf tubes (one per flask). After adding 1ml of chloroform, the samples are vortexed, centrifuged and dried as per Section 2.6.4 to detect L- and D-glutamate via GC-MS analysis.

2.8 RNA extraction

In preparation for gene expression analysis via RT-qPCR, RNA was extracted from cells using the QIAGEN RNeasy® Mini Kit (cat. Nos. 74104 and 74106) protocol.

2.8.1 Removal of DNA

To prepare DNA-free RNA before RT-qPCR, DNAase I (an endonuclease that digests single or double-stranded DNA) is added to the RNA preparation. For each RNA sample, 1 unit of DNAse I (RNase-free, #EN0521 Thermo Scientific) is combined with 1ug of RNA, 1µl of 10X reaction buffer with Magnesium Chloride and Diethylpyrocarbonate (DEPC)-treated Water to RNase-free tubes. The tubes are then incubated at 37ºC for 30 minutes. 1µM of 50mM Ethylenediaminetetraacetic acid (EDTA) is added to each tube and incubated at 65ºC for 10 minutes. The prepared RNA is then quantified before being stored in a -20ºC freezer and is used as a template for reverse transcription to cDNA.

2.8.2 RNA quantification

RNA quantification was performed using the Qubit® RNA HS (High Sensitivity) Assay Kits, which consist of a Qubit® HS reagent, buffer and two standards. The working solution was prepared by diluting the HS reagent at 1:200 in the buffer. 10µl of each standard is added into 190µl of working solution in 0.5ml clear PCR tubes to make up a final volume of 200µl per tube. 2µl of each RNA sample is added into 198µl working solution. All assay tubes are vortexed for two seconds and then incubated at room temperature for 2 minutes. The Qubit™ 4 Fluorometer is used to read the standards and samples. The standards are first read using the Qubit™ 4 Fluorometer to calibrate the assay, followed by the RNA samples and their concentrations are recorded.

2.9 Reverse transcription

Reverse transcription (RT) to obtain cDNA from the RNA samples was performed using the Add (Catalogue number 4387406, Appliedbiosystems, ThermoFisher Scientific). The kit components were thawed on ice. Once thawed, 10.0µl of 2X RT Buffer Mix and 1.0µl of 20X RT Enzyme Mix were combined to create a RT reaction mix which was then aliquoted to each sample tube containing up to 9µl of RNA. The amount of RNA added to each RT reaction is normalised to 1µg. The necessary volume of nuclease-free water was added to each RT reaction to make the total volume 20µl. Reaction tubes were pulse centrifuged to remove any air bubbles. The tubes were placed on ice. RT reactions were then transferred to a thermal cycler for incubation at 37ºC for 1 hour and 95ºC for 5 minutes. cDNA was stored at 4ºC for short-term and -20 for longer-term storage.

2.10 Real-time PCR (RT-qPCR)

TaqMan® Assays were obtained for four genes *DGLUCY, RSP6* (second target gene), *RPL32* and *PGK1* (housekeeping genes). All TaqMan® assays are for mouse genes, the housekeeping genes are from inventoried assays, and *DGLUCY* target genes are from custom-designed assays.

TaqMan® Fast Advanced Master Mix, TaqMan® Assay (20X) and cDNA samples were thawed on ice. The TaqMan® Assay (20X) and cDNA samples were then vortexed and centrifuged briefly. Individual PCR reactions contained 5µl of TaqMan® Fast Advanced Master Mix (2X), 0.5µl of TaqMan® Assay mix (20X) and 2.5µl of nuclease-free water per reaction. For ease of pipetting, master mixes were made up based on multiplying the volumes of the reagents required by 1+ the number of samples to be tested. The reaction was then vortexed and

centrifuged briefly to bring the mixture to the bottom and remove any air bubbles. The PCR reaction plate was then prepared by transferring the 8µl of PCR reaction to each well of the reaction plate containing 2µl of cDNA sample. Each cDNA was tested in duplicate for each gene, and water as the No Template Control (NTC) was prepared for each gene. The plate was sealed with an optical adhesive film and centrifuged briefly.

Real-time PCR was then performed using the QuantStudio™ 5 Real-Time PCR instrument using the 'Fast' run mode and FAM assay with the following protocol: Uracil-N-glycosylase (UNG) incubation at 50ºC (2 minutes), polymerase activation at 95ºC (2 minutes), denaturation at 95ºC (1 second), annealing at 60ºC (20 seconds). After the PCR, the data was analysed using the Design and Analysis Software v1.5.2 (ThermoFisher) to determine the threshold cycles (Ct) for the amplification curves. The Ct values represent the number of cycles required for the fluorescent signal (generated from the PCR reaction) to cross a background threshold level. They are inversely related to the amount of target gene in the sample analysed[189].

2.11 DNA extraction

Following the manufacturer's instructions, DNA was extracted from 500µl of whole blood using the Promega Maxwell RSC Whole Blood DNA kit and Promega Maxwell system (Maxwell ® RSC Instrument). Eluted DNAs were transferred to deep well plates (Fisher Scientific UK X50 Deep well Storage Plate 96 round wells, 1.2ml) and stored in the -20 and -80°C freezers.

To extract DNA from the saliva samples, the samples are first incubated at 50°C for 2 hours before storage in the freezer. Before DNA extraction, 20µl of prepIT.L2P (DNA Genotek) is added to 500µl of saliva sample and vortexed for a few seconds. The sample is incubated on ice for 10 minutes and centrifuged at 15000 x g for 5 minutes. The clear supernatant is

transferred into a new microcentrifuge tube and the remaining pellet containing impurities is discarded. 600µl of 100% ethanol is added to the supernatant and mixed gently 10 times. The sample is left at room temperate for 10 minutes for the DNA to precipitate fully. The sample is then centrifuged at 15000 x g for 2 minutes. The supernatant is removed and discarded without disturbing the pellet. An ethanol wash with 250µl of 70% ethanol is then added, with the ethanol entirely removed after 1 minute. 100µl of elution buffer is added to dissolve the DNA pellet and vortexed for five seconds. To allow rehydration of the DNA, the sample is incubated at room temperature and vortexed before storage and quantification.

2.12 DNA quantification

DNA was quantified using the Qubit™ dsDNA BR (Broad Range) assay kit. I prepared a 1 in 200 dilution of the Qubit™ BR reagent in Qubit™ buffer to produce 200µl of working solution for each standard and DNA sample. I set up two assay tubes as the standards (Standard 1 and 2). I prepared the assay tubes according to **Table 2.7.**

Table 2.7 Contents of assay tubes for DNA quantification.

	Standard assay tubes	**DNA Sample assay tube**
Volume of working solution (µl)	190	198
Volume of standard (µl)	10	-
Volume of DNA sample (µl)	-	2
Total volume per assay tube (µl)	200	200

Clear thin-wall 0.5ml PCR assay tubes were used. Abbreviations: Vol- volume

All the assay tubes were vortexed for two seconds and then centrifuged briefly at 5000rpm for 10 seconds. The tubes are then incubated for 2 minutes at room temperature. I first inserted Standard 1 and 2 in the Qubit™ 4 Fluorometer to calibrate the assay. Following this,

I then inserted each DNA sample assay tube in the fluorometer and recorded the DNA concentration readings of each sample in units ng/µl.

2.13 Whole Genome tagging SNP arrays

Genomic DNA from whole blood or saliva samples was used for genotyping. In the discovery cohort of the ICI toxicity project, only samples from patients who had toxicity data collected for at least three months were selected for genotyping. 250ng of DNA was normalised to 50ng/µl based on a minimum volume of 15µl per DNA sample. The normalised samples were plated in a 96-well plate and stored at -20°C, ready for shipment to our collaborator in the Spanish National Cancer Research Centre (CNIO) Human Genotyping Core Unit in Spain. The samples were genotyped using Illumina's Global Screening Array (GSA) v3.0, a whole genome tagging SNP array containing approximately 700,000 markers and some rare variants.

2.14 Quality controls steps applied to genotyping data

Genome Studio SNP genotyping module was used to perform genotype calling and clustering of markers based on build hg19/GRCh37. Cluster separation score threshold of 0.2 was used to identify good quality SNPs. GenCall score (how well each sample fits in the cluster) threshold of 0.15 and GenTrain score (how well genome studio was able to identify group of alleles) threshold of 0.4 were applied. I sorted the SNPs by GenTrain score and manually reviewed SNPS with the poorest clustering. Samples with call rates of <0.97 were excluded. Data generated after the application of these filters was exported in PLINK format. The following quality control (QC) steps were performed using PLINKv1.9 to identify:

1. Individuals or markers with elevated missing data rates. A threshold rate of 3% was applied for exclusion.
2. Duplicated/related individuals. Markers in linkage disequilibrium (LD) were identified through LD pruning (r^2 threshold 0.2). Long LD regions (including MHC loci) were excluded to decrease the computational complexity used to create the identity-by-state (IBS) matrix. Pair-wise IBS were generated for pairs of patients included in the analysis. Pairs of samples with high Pihat scores >0.05 indicating 4th-degree or higher relatives were identified, and patients with limited toxicity data available were excluded.
3. Individuals of divergent ancestry. ICI GENETICS SNP data was merged with the HapMapv3 dataset containing four ethnic populations. Only SNPs common to ICI GENETICS and HapMap were retained for principal component analyses (PCA) conducted using Flashpca 2.0. Plots of the first 10 PCs plotted against one another were generated using 'R' package ggplot2 v3.3.3.9000. Repeat PCA was performed without the HapMap population and without outlier samples. Outliers from PC1 and PC2 (6 standard deviations (SD) from the mean)) are excluded. I also checked whether any of the PCs were significantly associated with case-control status using ANOVA statistical testing in 'R studio' and did not find any significant association reaching P <0.05.

2.15 Phasing and Imputation

Variants that remained post QC that were not genotyped from the GSA were imputed using to infer and predict unobserved genotypes. Before imputation, phasing of genotype data from these remaining variants was performed using SHAPEIT (v2.r900)[190] to estimate the

haplotypes for imputation. I first removed any duplicate variants using PLINK v1.90b6.17 and flipped negative strand files based on the GSA-24v3-0 A2-b37 reference. To ensure the compatibility of the GWAS dataset with the reference panel of haplotypes, I checked strand alignment using SHAPEIT based on whole genome reference panels: 1000 Genomes phase 3 and UK10K merged reference panel containing 3871 samples and 91,414,192 markers (work courtesy of PhD student Laura Chegwidden) and Haplotype reference consortium (HRC)[164] containing 1227 samples and 40,405,530 markers.

I excluded mismatching SNPs before proceeding with phasing in SHAPEIT. Out of the 458743 SNPs genotyped variants inputted for phasing , 451117 variants remained after the checking and phasing steps described above. I next performed imputation of these variants based on the same reference panels using IMPUTE2 v2.3.2[164]. Following imputation, ~87 million variants and 654 individual samples were available for analysis. HLA classical alleles were imputed from the Type 1 Diabetes Genetics (T1DG) cohort using SNP2HLAv1.0.3[191].

2.16 Analysis of whole genome sequencing data

1µg of DNA from 8 patients who had developed multiple ≥grade 3 irAE in different organs was sent for 30x whole genome sequencing with Novogene. Raw sample demultiplexed FASTQ files were downloaded. Each sample had between one and three pairs of FASTQ files depending on whether the samples were sequenced in a single lane or across multiple lanes to achieve the desired average coverage of 30x.

The initial QC steps were performed on the sequencing data received for variant calling and annotation by my supervisor Claire Palles:

1) Pairs of FASTQ files were mapped to build GRCh38 using BWA-MEM[192].

2) BAM files generated were sorted using SAMTOOLSv1.12, and Picardv2.21.1 was used to add read groups. Duplicate reads in the BAM files were marked using Picard, and multiple BAM files from the same sample were combined. BAM files were then processed for compatibility with GATK Haplotype caller according to GATK Best Practices v4.1.0.

3) NM, MD and UQ tags were calculated by comparing with the reference using SetNmMdandUqTags with PICARD

   Base Recalibrator tool was run to build a model of covariation from the input data and a set of known variants from resources of known sites of variations (Broad 1000G) to generate a recalibration table. Base Quality Score Recalibration (BQSR) was applied to adjust the base quality scores based on the model and detect systematic errors from the sequencing machine during base call estimation

Genomic variant call format (GVCF) was then generated for each sample using HaplotypeCaller for variant calling. Joint calling across the samples was performed using GenotypeGVCFs. Single nucleotide variants (SNVs), insertion and deletions (indels) were called and then calibrated separately using VariantRecalibrator and Broad resources for SNPs (HapMap, Omni 2.5M SNP chip array, Homo Sapiens Assembly 38, 1000G Phase 1 SNPs, Mills and 1000G Gold standards) to assign a calibrated probability to each variant call in the set. ApplyVQSR was then performed on the indels and SNPs separately to apply a score cut-off for filtering variants based on a recalibration table produced by VariantRecalibrator. A filter level of 99.5 and 99 was used for SNPs and indels respectively. BCFtools v1.10.2 was then used to merge the indel and SNV calls, and only variants that passed the set filters were retained for analysis (PASS variants). LeftAlignAndTrimVariants were used to left-align variants before annotation. Multi-allelic variants were split into biallelic variants without trimming alleles and

stored in original AC, AF and AN values. VCFtools v0.1.16 was used to remove variants with missing information in any of the eight samples.

Variant annotation was performed for the PASS, left aligned, decomposed, and no missing variant calls using VEP96. VEP filters were applied to detect rare variants with allele frequencies <1% in non-Finnish Europeans or absent from the gnomAD database, which contains large-scale sequencing data from international coalitions[193]. **Figure 2.3** summarises the main steps taken to process the WGS data.

I applied other VEP filters to detect rare variants with possible Loss-Of-Function (LOF) consequences (splice acceptor, splice donor, stop gained or frameshift variants) or damaging/deleterious missense variants (SIFT ≤0.05 and Polyphen >0.908 scores). I also searched for variants mapping to a panel of any 395 candidate genes (associated with immuno-modulatory traits or nearby a locus from the discovery GWAS I conducted of ≥grade 3 irAE). This panel of 395 genes is listed in **Appendix 3**. I also searched for variants listed as pathogenic or likely pathogenic in the ClinVar database.

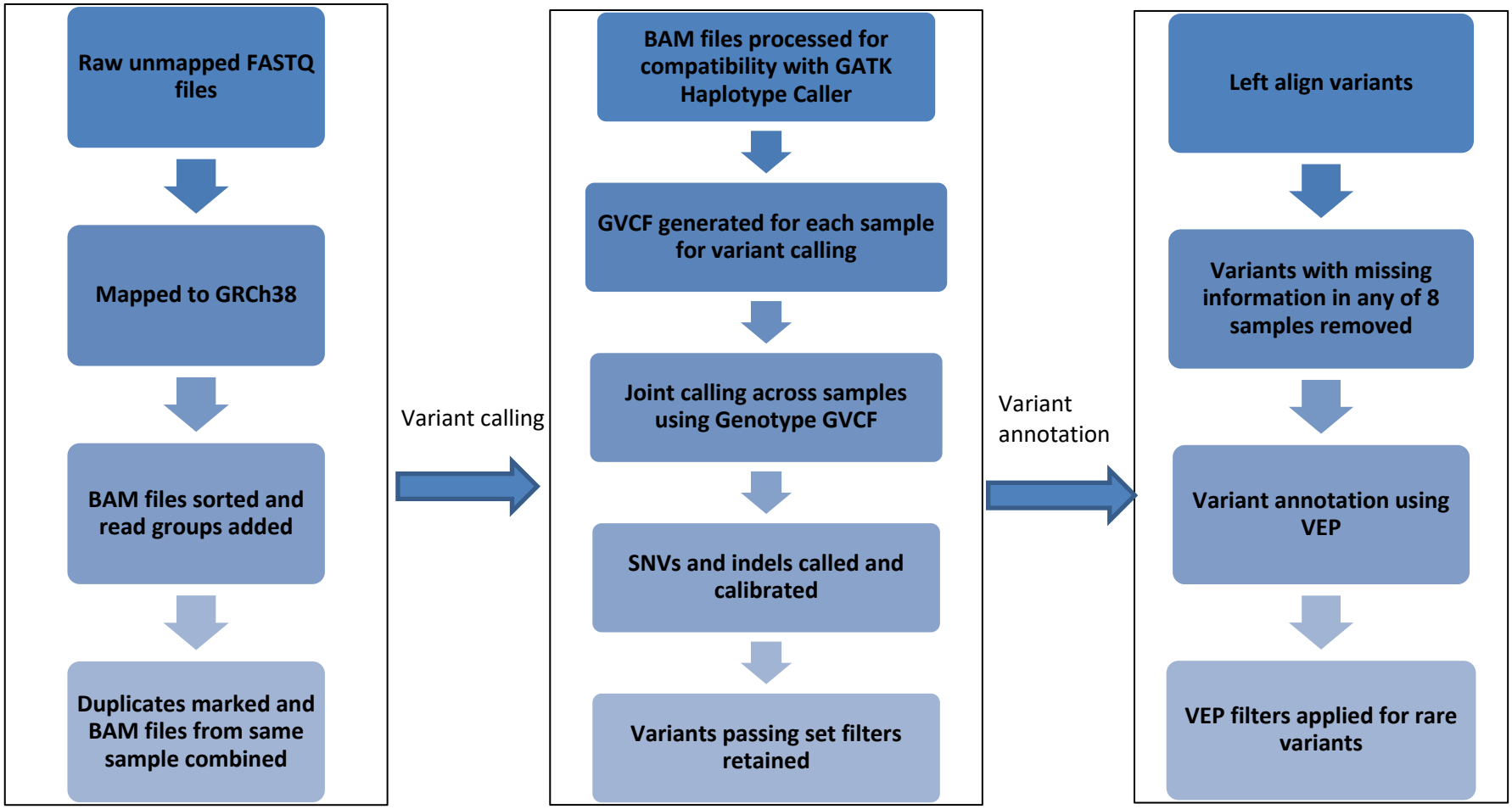


Figure 2.3 Summary of the pipeline used to process WGS data (work courtesy of Dr Claire Palles)

The main steps from initial receipt of raw sequencing data to variant annotation are presented. Abbreviations: GVCF: Genomic variant call format, SNVs- single nucleotide variants, indels- insertions and deletions, VEP- Ensembl Variant Effect Predictor software

CHAPTER 3: IDENTIFICATION OF GENETIC FACTORS PREDISPOSING TO FLUOROPYRIMIDINE-INDUCED TOXICITY

3.1 Introduction

Genetic variants have been shown to play a role in the development of fluoropyrimidine related toxicities. The protein dihydropyridine dehydrogenase (DPD) is the rate-limiting enzyme in the 5-FU metabolism pathway. Four variants in the *DPYD* gene encoding DPD have been validated as associated with ≥grade 3 5-FU toxicities[89]. These "no function" alleles (NM_000110.3:c.1905+1G>A (rs3918290) and NM_000110.3:c.1679T>G (rs55886062)) and "decreased function" alleles (NM_000110.3:c.2846T>A (rs67376798) and NM_000110.3:c.1129-5923C>G (rs56038477)) are being screened for in programmes in the UK before administration of 5-FU chemotherapy to guide treatment dosing and modifications[89]. Although *DPYD* variants are key predictors of 5-FU toxicity, they are thought to contribute to only a third of ≥grade 3 cases of 5-FU toxicity[194]. Current research has noted disparities between the expressions of *DPYD* mRNA, DPD protein and DPD enzymatic activity, suggesting other potential mechanisms could regulate DPD expression post-transcription[110].

Other non-genetic determinants of fluoropyrimidine-induced toxicity have also been studied. Branda et al. have found that 5-FU toxicity was significantly greater in rats that were folate-deficient compared to rats that were repleted or supplemented with folate[195]. The metabolic activity of DPD from peripheral blood mononuclear cells (PBMC) can also be evaluated by measuring the level of dihydrouracil, carbamyl-B-alanine and B-alanine produced from patient's lymphocytes when exposed to uracil, however this can be difficult to implement in clinical practice[196]. Endogenous uracil is metabolised into dihydrouracil by DPD in the liver and uracil levels can increase in DPD deficiency. Rising uracil plasma concentration was found to correlate with increasing severity of toxicity[197]. However, the relationship between uracil levels and DPD activity is still not clear[196], hence more studies

are required to establish the applicability of plasma uracil concentration in determining risk of fluoropyrimidine toxicity.

3.1.1 Genetic determinants of capecitabine-induced toxicity

Capecitabine is an oral pro-drug that is activated to 5-FU as per Figure 1.2 and used widely in treating various cancer types, including gastrointestinal and breast cancer. Like intravenous (IV) 5-FU, capecitabine can be given as monotherapy or combined with other chemotherapies or targeted agents. Both IV 5-FU and capecitabine commonly induce gastrointestinal toxicities, with diarrhoea, nausea and vomiting being comparable between the two routes of drug administration[198]. Diarrhoea and nausea occur early within a week of treatment[89]. However, hand-foot syndrome (HFS) is significantly more common following treatment with capecitabine compared to 5-FU (Relative Risk (RR) 3.45, $P<0.00001$), whereas the risk of ≥grade 3 mucositis and neutropenia is lower with capecitabine than with 5-FU[198]. Capecitabine can be given at a dose of 1000-1250mg/m^2 twice daily as a single agent or at a lower dose in combination regimes, but dose reductions are common[199]. Renal impairment involving a 50% decrease in creatinine clearance has been linked with a reduction in clearance of the toxic metabolite fluoro-beta-alanine (FBAL), which is correlated with an increased risk of severe diarrhoea[200]. Although several patient and treatment factors can influence the development of treatment-related toxicities, genetic variants have also been widely studied for their ability to predict toxicity occurrence. Studies to date have reported associations of capecitabine toxicity with variants in 5-FU metabolising enzymes and thymidylate synthase such as *TYMS*, *DPYD*, *CDA, CES2* and MTHFR (Chapter 1: Table 1.2)[8]. However, reported findings have been inconsistent between studies, likely due to small sample sizes, heterogeneity in sample populations and different drug regimens used[8].

3.1.2 The role of mir27A in predicting risk of early-onset fluorouracil toxicity

Mir27a and mir27b microRNAs play a role in regulating DPD expression in the liver, a key metabolism site of 5-FU chemotherapy. In silico modelling has shown that they bind to two conserved sites in *DPYD* post-transcriptionally[110]. Overexpression of mir27a and mir27b are associated with reduced DPD expression by regulating the targeting of RNA-induced silencing complex (RISC) complexes to *DPYD*, which leads to the degradation of its mRNA[110]. *MIR27A* variant rs895819, located within the hsa-mir-27a hairpin coding region, has been tested for its association with DPD expression[110]. Rs895819 T>C is a common variant with a MAF of 33% in non-Finnish Europeans, 35% in South Asians and 50% in the African/African American population[193]. Expression of mir-27A in lymphoblastoid cell lines homozygous for the variant 'C' allele was significantly higher (by ~50%) compared to lines that were homozygous for the wild type 'T' allele, with heterozygous cell lines showing intermediate levels of expression[110]. A multivariate analysis showed that human carriers of rs895819 had significantly less DPD enzyme activity in peripheral blood mononuclear cells than non-carriers[110]. Another less common *MIR27A* variant, rs11671784 G>A with MAF of 2% in the non-Finnish European, 0.5% in South Asian and 0.3% in African/African American population[193] has also been found to affect miR-27a expression[115]. Carriers of the 'A' allele have been associated with lower miR-27a expression and can also reduce the upregulating effect of rs895819[201]. The same study reported that the three genotypes of rs11671784 did not find any significant difference expression in total pri- and pre-mir27a levels, indicating that the downregulating effect on mir27a expression may be due to a block from pre-mir27a to its mature form[201]. Yang et al. similarly reported that rs11671784 G/A did not affect mir27a transcription[202].

Amstutz et al. assessed 514 patients receiving 5-FU based chemotherapy and found an association between the common *MIR27A* variant rs895819 and the risk of early-onset ≥grade 3 toxicities within the first two cycles[111]. A study by Meulendijks et al. involved 1427 patients on capecitabine-based chemotherapy and found an association between rs895819 and ≥grade 3 toxicity during the first treatment cycle[115]. However, both studies observed that the association between rs895819 and early-onset toxicity was only significant in patients who also carried a *DPYD* "no function" or "low function" variant allele (7% and 10% of total patients included in Amstutz and Meulendijks et al's studies respectively) and not in patients with wild type *DPYD* [111,115]. This observation was further strengthened by a meta-analysis of the two studies demonstrating a significant association between rs895819 and 5-FU toxicity in carriers of *DPYD* "no function" or "low function" variants (OR 5.4, 95% CI 1.8-15.7, p=0.002)[115]. However, the validity and reproducibility of this finding may be limited by the small number of patients with *MIR27A* rs895819 variants who are also *DPYD* carriers from both Amstutz and Meulendijks studies (N=19/514 (3.7%) and 96/1592 (6%) respectively) [111,115]. The association between rs11671784, a rare *MIR27A* variant, and 5-FU toxicity was also tested by Meulendijks et al. but was not significant (p=0.777)[115].

Previous analyses conducted by my lab group did not identify an association between rs895819 and global capecitabine-induced ≥grade 3 toxicity in the QUASAR2 trial dataset but did identify significant associations between this SNP and risk of ≥grade 3 diarrhoea (occurring across all cycles of treatment. There was also suggestive evidence of a protective effect on haematological toxicity **(Table 3.1)**.

Table 3.1 Association analysis of rs895819 with capecitabine toxicity across all-treatment cycles in QUASAR2 based on imputed data, N=930 patients (previous work conducted by Palles group)

All cycles	OR	95% CI	P value
Diarrhoea	1.55	1.11-2.15	0.009
Haem	0.53	0.23-1.24	0.145
HFS	0.88	0.68-1.13	0.311
Mucositis	1.15	0.48-2.76	0.751
Vomiting	0.92	0.40-2.11	0.852
Global	1.01	0.81-1.26	0.906
GI	1.48	1.07-2.03	0.016

Summary statistics of the different toxicity variables tested based on imputed data using the QUASAR2 dataset. This analyses was previously conducted by my lab group (Palles group) independently of the analyses from my thesis. Abbreviations: CI- confidence interval, GI- gastrointestinal, Haem- haematological, HFS- hand-foot-syndrome, OR- odds ratio

3.1.3 Genetic determinants of maximum grade all cycle capecitabine-induced toxicity in the QUASAR2 trial

Candidate and pathway analysis of genetic variants associated with maximum grade all cycle capecitabine-induced toxicity using QUASAR2 samples have been published[12,101]. A genome-wide analysis of genetic markers correlated with any-cycle ≥grade 3 capecitabine-induced toxicity had also been performed by a previous member of the Palles lab (Marie Lucas) but had not been published due to the lack of a validation cohort. Analysis of toxicity events occurring in the first two treatment cycles has not previously been investigated in QUASAR2. It can be argued that these early events may have a genetic basis as demonstrated by the influence of *MIR27A* variants on early-onset fluoropyrimidine-induced toxicity[111], and they are also expected to lead to early dose reductions or cessation of treatment, potentially resulting in limited anti-cancer benefits to the patient.

3.1.4 Chapter aims

To analyse the impact of previously reported and novel variants associated with early-onset capecitabine toxicity, this chapter had the following aims:

1. Identify genetic variants predictive of CTCAE ≥grade 3 capecitabine toxicity in the QUASAR2 trial and from meta-analyses of the QUASAR2 and COIN trials.
2. Validate the association between two *MIR27A* SNPs (rs895819T>C and rs11671784G>A) and early ≥grade 3 fluoropyrimidine-related toxicity (cycle 1 and cycles 1 and 2) using the QUASAR2 dataset.
3. Evaluate whether the *DPYD* low and no function variants previously reported to be associated with fluoropyrimidine toxicity were associated with early ≥grade 3 toxicity in QUASAR2 and whether there was an interaction between *DYPD* status at these alleles and *MIR27A* rs895819 genotype.
4. Determine the relationship between *MIR27A* SNPs and ≥grade 3 early fluoropyrimidine-related toxicity from meta-analyses of the QUASAR2 trial and a Dutch cohort of patients enrolled in a prospective, multicenter study of *DPYD*2A* genotype-guided dosing of fluoropyrimidines.

3.2 Results

3.2.1 Patient cohort

The number of patients from the QUASAR2 trial with early-onset CTCAE grade 0 to 2 and grade 3 to 4 capecitabine-induced toxicities during their first two treatment cycles and across all treatment cycles is shown in **Table 3.2**. Whilst the number of grade 3 to 4 individual toxicities during the first two cycles is modest in QUASAR2 (0.9-8.3%), 15.3% experienced one or more early ≥grade 3 events of any type (global toxicity variable). Toxicity outcome variables with

less than 10 grade 3-4 cases across cycles 1 and 2 were excluded due to the low power given the low number of cases. Assuming ~30% of patients are at risk of developing ≥grade 3 fluoropyrimidine toxicity, the analysis will be adequately powered (>80%) to detect common variants with minor allele frequencies (MAF) of 30% at an odds ratio (OR) of 1.4 and detect rare variants (MAF 10%) at OR 1.6.

Table 3.2 Number of patients with capecitabine-induced toxicities in QUASAR2

Total N=930	**Cycle 1**			**Cycles 1 & 2**			**All cycles**		
Toxicity	**Grade 0-2 (n, %)**	**Grade 3-4 (n, %)**	**NA (n, %)**	**Grade 0-2 (n, %)**	**Grade 3-4 (n, %)**	**NA (n, %)**	**Grade 0-2 (n, %)**	**Grade 3-4 (n, %)**	**NA (n, %)**
Diarrhoea	896 (96.3)	27 (2.9)	7 (0.8)	864 (92.9)	64 (6.9)	2 (0.2)	829 (89.1)	99 (10.6)	2 (0.2)
Haem	917 (98.6)	9 (1.0)	4 (0.4)	915 (98.4)	12 (1.3)	3 (0.3)	908 (97.6)	19 (2.0)	3 (0.3)
HFS	903 (97.1)	14 (1.5)	13 (1.4)	850 (91.4)	77 (8.3)	3 (0.3)	716 (77.0)	211 (22.7)	3 (0.3)
Mucositis	919 (98.8)	8 (0.9)	3 (0.3)	918 (98.7)	10 (1.1)	2 (0.2)	915 (98.4)	13 (1.4)	2 (0.2)
Vomiting	919 (98.8)	3 (0.3)	8 (0.9)	918 (98.7)	8 (0.9)	4 (0.4)	911 (98.0)	15 (1.6)	4 (0.4)
Global	882 (94.8)	46 (5.0)	2 (0.2)	787 (84.6)	142 (15.3)	1 (0.1)	621 (66.8)	308 (33.1)	1 (0.1)

The frequency of toxicity cases during the first two cycles and across all cycles of treatments according to CTCAE grading is shown. Missing data from each toxicity variable is coded as NA. Haematological toxicity variable is defined as the presence of neutropenia and/or thrombocytopenia; mucositis is defined as mucositis and/or stomatitis, and global is defined as any toxicity type. Abbreviations: Haem- haematological, HFS- hand-foot-syndrome

3.2.2 Analysis of previously reported variants for association with early cycle capecitabine induced toxicity in QUASAR2

Six variants have previously been reported to be associated with early onset 5-FU toxicity [107,111,203]. These include the four “no function” or ”low function” *DPYD* risk variants included in the Clinical Pharmacogenetics Implementation Consortium (CPIC®) guidelines, one additional *DPYD* variant (c.1601G>A) and a *MIR27A* variant (rs895819A>G). I analysed the association between these variants and early-onset toxicity in the QUASAR2 dataset **(Table**

3.3). 22/142 (15%) of the patients experiencing ≥grade 3 events in the first two cycles are carriers of one of the four *DPYD* alleles listed in CPIC®. *DPYD* risk variants c.2846A>T and rs56038477 (tags HapB3) are significantly associated with the presence of any toxicity (global toxicity) from the first two cycles (P=0.001 and 0.0002 respectively). c.1905+1G>A also showed suggestive evidence of an association with early cycle toxicity but did not reach statistical significance. It was not possible to assess c.1679T>G by itself as there is only a single carrier of this risk allele in QUASAR2. The associated confidence intervals from the summary statistics of the four *DPYD* risk variants are also noticeably wide, likely reflected by the low MAFs of these variants in QUASAR2 (0.05%-1.9%) and low number of DPYD risk carriers that have a ≥grade 3 event (n=22). *DPYD* variant c.1601G>A, reported in the literature to be linked with early toxicity but not included in the screening test, was found to have a significant association with global toxicity in our analysis (P=0.016). No statistically significant increase in risk was observed between the previously reported common *MIR27A* variant rs895819A>G and early-onset global toxicity (OR=1.13, P=0.4). T

Table 3.3 Association between previously reported variants with early onset global toxicity during the first two treatment cycles in QUASAR2

Gene	Variant	Capecitabine induced global toxicity in cycles 1 & 2		
		OR	**95% CI**	**P value**
DPYD	c.2846A>T/rs67376798*	8.76	0.34-223.96	0.001
	c.1905+1G>A/rs3918290*	3.93	0.86-17.93	0.100
	c.1679T>G/rs55886062*	NA	NA	NA
	c.1129–5923C>G/HapB3/rs56038477*	3.32	1.64-6.72	0.002
	Any CPIC toxicity risk allele	4.05	2.28-7.17	5.4×10^{-6}
	c.1601G>A/rs1801158	2.53	1.24-5.14	0.016
MIR27A	rs895819A>G	1.13	0.85-1.51	0.400

Abbreviations: CI- confidence interval.**DPYD* variants included in the CPIC guidelines. NA- Unable to analyse due to insufficient cases.

3.2.3 Genome-wide association analysis of capecitabine toxicity in QUASAR2

I used SNPTESTv2.5.6 to perform single SNP association analysis using frequentist testing to identify SNPs predictive of early-onset and any cycle ≥grade 3 capecitabine-induced toxicities including diarrhoea, mucositis, haematological, hand-foot-syndrome and vomiting in QUASAR2. The maximum grade toxicity events in cycles 1 and 2 is being analysed for each trial participant for the early-onset analysis and across all 8 cycles for the any cycle analysis. I generated binary coded variables of the listed capecitabine toxicities using a case-control approach. Cases consisted of ≥grade 3 toxicity events coded as '1' and controls are grade 0-2 events coded as '0'. The global toxicity variable is coded as "1" when a ≥grade 3 event of any type of toxicities listed above is present and "0" if none of the toxicities were graded higher than grade 2. I applied the additive genetic model for each genome-wide association analysis of the individual toxicity variable analysed as this model is commonly used in case-control GWAS and performs well to detect variants that have additive or dominant heritance pattern[204]. The additive model counts the number of hypothesises that genotype MM (Major allele homozygote), Mm (heterozygote) and mm (minor allele homozygote) are associated with lowest, intermediate and highest risk respectively[205]. The two trial arms of QUASAR2 (Arm A: capecitabine treatment and Arm B: capecitabine and bevacizumab treatment) were initially analysed separately with SNPTESTv2.5.6. Instead of using the two arms of QUASAR2 as a discovery and validation set, meta-analysis of the two arms were conducted using METAv1.7 due to the lack of validation cohort at the time of analysis.

3.2.4 Top SNPs associated with early-onset toxicity using different methods to handle genotype uncertainty

The software used to perform association testing (SNPTEST v2.5.6) has several methods for handling genotype uncertainty, as described in methods chapter Section 2.3.2. The "score" method in SNPTEST uses the observed data likelihood (contribution of each genotype weighted by its imputation probability) in the association analysis[167]. Here the software attempts to maximise the likelihood by maximising the first and second derivates of the likelihood under the null hypothesis. The "score" method is suitable for associations where the log-likelihood is similar to a quadratic[167]. However, small sample population, rare risk variants, and high genotyping uncertainty from imputation weakens the likelihood and generates falsely low P-values[167].

3.2.5 Genome-wide association analysis of early-onset capecitabine toxicity

Using the "score" method approach to handle genotype uncertainty, 14 independent SNPs were identified to be associated with early-onset capecitabine-induced ≥grade 3 toxicities at genome-wide significance ($P<5x10^{-8}$) **(Table 3.4).**

Table 3.4 Top independent SNPs from GWAS of early-onset (cycles 1 and 2) ≥grade 3 capecitabine toxicity using the "score" approach to handle genotype uncertainty under the additive genetic model

Early Toxicity	rsIDs	Chr	Pos	Risk allele	Freq	INFO	OR	Lower CI	Upper CI	P value	Consequences	Mapped gene/nearest genes
Haem	rs698607	2	191052117	G	0.07	0.84	742	113	4861	5.49E-12	intronic	C2orf88
Haem	rs115143613	3	187661895	T	0.06	0.93	562	66	4734	5.68E-09	intergenic	BCL6(dist=198382),LINC01991(dist=14668)
Haem	rs147827021*	7	121686924	T	0.07	0.96	266	39	1783	8.77E-09	intronic	PTPRZ1
Haem	rs199983532	10	133335072	A	0.06	0.94	223	33	1500	2.64E-08	intergenic	TCERG1L (dist=225088),LINC01164 (dist=269662)
Haem	rs1327306	13	27670283	T	0.07	0.99	111.96	21.31	588.24	2.49E-08	intronic	USP12
Haem	rs4792825	17	43919681	G	0.1	0.97	74	16.47	331.85	1.96E-08	intergenic	LINC02210-CRHR1(dist=6487),MAPT-AS1(dist=1041)
Haem	rs78361592	20	56132320	G	0.05	0.93	426	53	3441	1.31E-08	intergenic	CTCFL(dist=31612),PCK1 (dist=3817)
Mucositis	rs2929329	1	218769034	T	0.15	0.94	58.45	14.11	242.17	2.04E-08	ncRNA intronic	MIR548F3
Mucositis	rs35741121	6	90109696	A	0.09	0.93	465	72	3004	1.11E-10	intronic	RRAGD
Mucositis	rs56329496	6	135530039	A	0.1	0.86	275	45	1670	1.05E-09	intronic	MYB
Mucositis	rs117810108	7	130434431	G	0.05	0.81	2202	205	23677	2.13E-10	intergenic	KLF14 (dist=15571), MIR29A (dist=127075)
Mucositis	rs149994224*	10	85622866	A	0.12	0.92	181	32.33	1010.32	3.23E-09	intergenic	LOC105378397(dist=186021), GHITM(dist=276319)
Mucositis	rs11024884	11	18950092	G	0.09	0.92	349	43.72	2787.34	3.31E-08	intergenic	PTPN5(dist=135824),MRGPRX1(dist=5268)
Mucositis	rs111378975*	12	120838305	T	0.05	0.92	9598	812	113409	3.40E-13	intergenic	MSI1(dist=31322),COX6A1(dist=37588)

Conditional and joint ('COJO') analysis was used to identify top SNPs associated with each toxicity variable (CTCAE ≥grade 3 events) that reached GWAS significance ($P<5x10^{-8}$) after filtering for MAF>1% in cases and controls, presence in both trial arms (n>1), P heterogeneity >0.05 and adjusting for covariates age and sex. The risk allele refers to the allele associated with toxicity risk. Abbreviations: Chr -chromosome, CI- confidence interval, Freq -frequency of risk allele, Haem- haematological, INFO- info score, OR -odds ratio, Pos- genomic position (build GRCh37), ncRNA- non-coding RNA, UTR- untranslated region. *SNP associated with corresponding toxicity variable in both early cycles and all cycles analysis under the "score" model in SNPTEST .

As seen in **Table 3.4,** the ORs for the SNPs reaching genome-wide significance are implausibly large. For example, ~5% of patients who are carriers of the risk allele 'G' of rs117810108 developed ≥grade 3 mucositis. Due to the uncertainty around the validity of the "score" model, I used the "expected" method as an alternative approach for handling genotype uncertainty. This method uses the expected allele counts to test for association with a binary trait using a standard logistic regression model or a quantitative trait using the linear regression model[167]. The "expected" approach provides a good approximation of genotype uncertainty for analyses where the risk allele's effect size is small, which is found for most common variants in GWAS[167]. By using the "expected" model for the GWAS of early ≥grade 3 capecitabine toxicity, the outputs of the analyses showed less inflated association statistics **(Table 3.5).** Only one new SNP (rs181841887) was identified in the "expected" model to be associated with early-onset ≥grade 3 HFS at the GWAS significance threshold of $P<5x10^{-8}$. This SNP mapped to the *GALNT17* gene. 14 new SNPs showed suggestive associations at $P<5x10^{-7}$ **(Table 3.5).**

Table 3.5 Top independent SNPs from GWAS of early-onset (cycles 1 and 2) ≥grade 3 capecitabine toxicity using the "expected" approach to handle genotype uncertainty under the additive genetic model

Early Toxicity	rsIDs	Chr	Pos	Risk allele	Freq	INFO	OR	Lower CI	Upper CI	P value	Consequences	Mapped gene /nearest genes
Diarrhoea	rs8048167	16	73570551	C	0.18	0.70	3.70	2.28	6.01	1.24E-07	Intergenic	LINC01568 (dist=115256), LOC101928035 (dist=655740)
Diarrhoea	rs10654492	21	21616413	C	0.32	0.64	3.44	2.16	5.49	2.00E-07	Intergenic	LINC01683 (dist=344344), LINC02573 (dist=12762)
Diarrhoea	rs35087079	18	57863787	G	0.02	0.72	13.21	4.97	35.06	2.22E-07	Intergenic	PMAIP1 (dist=292249), MC4R (dist=174777)
Diarrhoea	rs117686094	8	26875681	A	0.03	0.98	5.48	2.87	10.47	2.50E-07	Intergenic	ADRA1A (dist=150921), STMN4 (dist=217159)
Diarrhoea	rs117260063	7	31923305	T	0.04	0.83	6.57	3.21	13.44	2.58E-07	Intronic	PDE1C
Diarrhoea	rs12231444	12	124074365	T	0.03	0.95	5.90	2.99	11.63	3.09E-07	Intronic	TMED2
Diarrhoea	rs1491207661	3	61476710	TA	0.15	0.75	3.44	2.13	5.57	4.76E-07	Intergenic	FHIT (dist=239575), PTPRG (dist=70532)
Diarrhoea	rs11643168	16	85981482	T	0.07	0.72	4.80	2.61	8.85	4.82E-07	Intergenic	IRF8 (dist=25267), LINC01082 (dist=248305)
Vomiting	rs193272564	11	97684338	T	0.01	0.81	70.81	14.84	337.91	9.17E-08	Intergenic	LOC105369443(dist=1436424),CNTN5(dist=1207368)
Vomiting	rs140744481	13	70848486	A	0.02	0.94	51.55	11.17	237.93	4.37E-07	Intergenic	ATXN8OS(dist=134601),LINC00348(dist=740784)
Mucositis	rs76211390*	14	45617764	C	0.04	0.97	37.14	9.79	140.84	1.07E-07	Intronic	FANCM
Mucositis	rs79272399	14	99945412	C	0.02	0.91	41.56	9.88	174.88	3.71E-07	Intronic	SETD3
Global	rs71313929	22	19713353	G	0.61	0.61	2.67	1.83	3.90	3.74E-07	Intergenic	SEPT5-GP1BB (dist=1056), TBX1 (dist=30873)
HFS	rs181841887	7	71166314	A	0.03	0.95	6.409	3.33	12.33	2.68E-08	Intronic	GALNT17

Conditional and joint ('COJO') analysis was used to identify top SNPs associated with each toxicity variable (CTCAE ≥grade 3 events) that reached a P-value threshold of $<5x10^{-7}$ after filtering for MAF>1% in cases and controls, presence in both trial arms (n>1), P heterogeneity >0.05 and adjusting for covariates age and sex. The risk allele refers to the allele associated with toxicity risk. Abbreviations: Chr -chromosome, CI- confidence interval, Freq -frequency of risk allele, Haem- haematological, HFS- hand-foot-syndrome, INFO- info score, OR -odds ratio, Pos- genomic position (build GRCh37)**.** *SNP associated with the same toxicity in the all cycles analysis under the "expected" model in SNPTEST.

3.2.5.1 Comparison of the method score and method expected approach

To assess the concordance of both methods in handling genotype uncertainty, **Table 3.6** compares the summary statistics of the top independent SNPs identified from the "score" method approach in the "expected" method analysis. None of the variants associated with early-onset toxicity under the "score" model retained GWAS significance in the "expected" model. Although the SNPs that were GWAS significant from the "score" method were not as strongly significant in the "expected" method analysis, they were still at an adequate level of significance ($P<1x10^{-5}$), suggesting potential associations with the corresponding toxicity variables analysed. SNPs identified by the "expected" method approach were used in the polygenic risk score assessment in Section 3.2.6.

Table 3.6 Summary statistics of the top SNPs from the score method approach in both the score and expected method analysis

					Method Score				Method Expected			
Early Toxicity (cycles 1 and 2)	RSIDs	Chr	Pos	Freq	OR	Lower CI	Upper CI	P value	OR	Lower CI	Upper CI	P value
Haem	rs698607	2	191052117	0.07	742	113	4861	5.49E-12	11.49	3.83	34.48	1.339E-05
Haem	rs115143613	3	187661895	0.06	562	66	4734	5.68E-09	9.09	3.10	26.70	5.885E-05
Haem	rs147827021	7	121686924	0.07	266	39	1783	8.77E-09	8.11	2.96	22.16	4.55E-05
Haem	rs199983532	10	133335072	0.06	223	33	1500	2.64E-08	9.35	3.18	27.44	4.73E-05
Haem	rs1327306	13	27670283	0.93	0.01	0.0017	0.05	2.49E-08	0.13	0.05	0.32	1.08E-05
Haem	rs4792825	17	43919681	0.1	74	16.47	331.85	1.96E-08	10.09	3.53	28.80	1.57E-05
Haem	rs78361592	20	56132320	0.05	426	53	3441	1.31E-08	10.93	3.45	34.64	4.82E-05
Mucositis	rs2929329	1	218769034	0.85	0.02	0.004	0.07	2.04E-08	0.12	0.05	0.33	2.71E-05
Mucositis	rs35741121	6	90109696	0.09	465	72	3004	1.11E-10	8.81	3.19	24.39	2.78E-05
Mucositis	rs56329496	6	135530039	0.1	275	45	1670	1.05E-09	11.62	3.70	36.49	2.66E-05
Mucositis	rs117810108	7	130434431	0.05	2202	205	23677	2.13E-10	10.19	3.47	29.89	2.38E-05
Mucositis	rs149994224	10	85622866	0.12	181	32.33	1010.32	3.23E-09	7.71	2.85	20.84	5.74E-05
Mucositis	rs11024884	11	18950092	0.09	349	43.72	2787.34	3.31E-08	9.46	2.94	30.42	1.64E-04
Mucositis	rs111378975	12	120838305	0.05	9598	812	113409	3.40E-13	13.90	4.59	42.04	3.15E-06

The effect sizes and P values of top SNPs associated with early ≥grade 3 capecitabine toxicity identified using the score model are compared with those generated when the expected method is used to account for genotype uncertainty. Abbreviations: Chr -chromosome, CI- confidence interval, Freq -frequency, Haem- haematological, OR- odds ratio, Pos- genomic position (build GRCh37)

3.2.6 Genome-wide association analysis of capecitabine toxicity across all cycles

A previous study by Rosmarin et al analysed the association of 36 candidate polymorphisms identified from the literature with global capecitabine toxicity in the QUASAR2 trial[101]. The authors found that the rare *DPYD* alleles 2846T>A and *2A and common *TYMS* polymorphisms 5'VNTR2R/3R and 3'UTR 6bp ins-del were associated with global ≥grade 3 toxicity with nominal significance (P<0.05)[101]. Given these promising findings, I performed a hypothesis-free GWAS using the same QUASAR2 dataset to identify other variants that could be predictive of ≥grade 3 toxicity across all cycles. The GWAS was conducted using the "expected" method because it was a more biologically plausible method for handling genotyping uncertainty. Using this method, 13 SNPs were associated with any-cycle capecitabine toxicity at suggestive significance with $P<5x10^{-7}$ **(Table 3.7).** One of these**,** rs76211390, associated with any-cycle mucositis also showed a suggestive significant association with early-cycle mucositis. This SNP maps to the *FANCM* gene, which is linked to breast cancer chemotherapy toxicity[206]. rs12211900 was the only marker to reach GWAS significance in the analysis of any-cycle toxicity. It was associated with all cycle diarrhoea (OR=2.88, 95% CI=1.97-4.19, $P=3.92x10^{-8}$) and maps to an intergenic region; nearest genes *LINC02520* and *MIR4462*.

Table 3.7 Top independent SNPs from GWAS of ≥grade 3 capecitabine toxicity across all treatment cycles using the "expected" approach to handle genotype uncertainty under the additive genetic model

Toxicity	RSIDs	Chr	Pos	Risk allele	Freq	INFO	OR	Lower CI	Upper CI	P value	Consequences	Mapped gene /nearest genes
Diarrhoea	rs12211900	6	37510161	C	0.11	0.95	2.88	1.97	4.19	3.92E-08	intergenic	LINC02520 (dist=6105), MIR4462 (dist=12980)
Diarrhoea	rs71379941	17	63589380	C	0.02	0.75	8.66	3.75	19.96	4.14E-07	intergenic	AXIN2 (dist=31614), CEP112 (dist=42278)
Haem	rs75217739	22	45577373	T	0.06	0.65	30.07	8.24	109.73	2.55E-07	ncRNA intronic	LOC105373064
Haem	rs139788123	12	18946760	A	0.03	0.71	21.37	6.59	69.23	3.32E-07	Intergenic	CAPZA3 (dist=54638), PLEKHA5 (dist=335866)
HFS	rs79986886	5	165441781	G	0.04	0.83	4.12	2.45	6.94	1.03E-07	intergenic	LOC102546299 (dist=1471792), LINC01947 (dist=890446)
HFS	rs116448748	1	97490443	A	0.03	0.97	4.57	2.55	8.16	3.00E-07	intergenic	PTBP2 (dist=209838), DPYD (dist=52857)
Mucositis	rs77911328	2	208627843	T	0.02	0.64	45.28	11.20	183.14	8.90E-08	UTR3	FZD5 (NM_003468:c.*3863G>A)
Mucositis	rs145225093	9	84407550	C	0.02	0.70	23.80	7.16	79.16	2.32E-07	intergenic	LOC101927502 (dist=15736), SPATA31D5P (dist=120802)
Mucositis	rs76211390*	14	45617764	C	0.04	0.97	19.31	6.21	60.09	3.20E-07	intronic	FANCM
Mucositis	rs117241924	8	137378694	A	0.02	0.85	42.87	10.07	182.53	3.68E-07	intergenic	KHDRBS3 (dist=718842), LINC02055 (dist=164347)
Vomiting	rs200676466	11	121702545	CT	0.02	0.76	39.34	10.47	147.84	5.41E-08	intergenic	SORL1 (dist=198074), MIR100HG (dist=196492)
Vomiting	rs4972833	2	173372371	A	0.02	0.69	25.395	7.54	85.53	1.78E-07	intergenic	ITGA6 (dist=1188), PDK1 (dist=48326)
Vomiting	rs148123087	2	213644924	C	0.02	0.85	22.763	6.89	75.19	2.96E-07	intergenic	ERBB4 (dist=241572), LINC01878 (dist=15131)

Conditional and joint ('COJO') analysis was used to identify top SNPs associated with each toxicity variable (CTCAE ≥grade 3 events) that reached a P-value threshold of $<5x10^{-7}$ after filtering for MAF>1% in cases and controls, presence in both trial arms (n>1), P heterogeneity >0.05 and adjusting for covariates age and sex. The risk allele refers to the allele associated with toxicity risk. Abbreviations: Chr -chromosome, CI- confidence interval, Freq -risk allele frequency, Haem- haematological, HFS- hand-foot-syndrome, INFO- info score, OR -odds ratio, Pos- genomic position (build GRCh37) *SNP associated with the same toxicity in the cycle 1 and 2 analysis under the "expected" model in SNPTEST.

3.2.7 Meta-analysis GWAS of early ≥grade 3 capecitabine toxicity

Despite much effort, I could not identify a large validation cohort with blood samples and cycle-by-cycle toxicity data from an independent set of stage II/III colorectal cancer (CRC) patients treated with capecitabine. Toxicity by cycle data was however available from a clinical trial in the advanced CRC patient setting: COIN (Chapter 2: Section 2.1.2). COIN was also analysed for associations between genetic markers and early-onset toxicity by Katie Watts, PhD student in Jeremy Cheadle's group at the University of Cardiff. Imputed variants from COIN were hard calls generated from PLINK and those with a genotype uncertainty score of <0.9 were excluded (methods chapter Section 2.3.3). Replication of the top hits (P value threshold $5x10^{-8}$) from each trial was sought in the other trial before looking for markers reaching genome-wide significance in a combined meta-analysis. **Table 3.8** shows the frequency of early-onset toxicities in COIN. The frequency of these adverse events in QUASAR2 is also shown for comparison.

Table 3.8 Number of early-onset toxicity cases in both QUASAR2 and COIN trial cohorts.

Toxicity	QUASAR2 (N=930)			COIN (N=998)		
	Grade 0-2 (n, %)	Grade 3-4 (n, %)	NA (n,%)	Grade 0-2 (n, %)	Grade 3-4 (n, %)	NA (n,%)
Diarrhoea	864 (92.9)	64 (6.9)	2 (0.2)	883 (88.5%)	115 (11.5%)	2 (0.2%)
Haematological (QUASAR2)/Neutropenia (COIN)	915 (98.4)	12 (1.3)	3 (0.3)	985 (98.7%)	10 (1%)	3 (0.3%)
Hand-foot syndrome	850 (91.4)	77 (8.3)	3 (0.3)	984 (98.6%)	11 (1.1.%)	3 (0.3%)
Mucositis	918 (98.7)	10 (1.1)	2 (0.2)	990 (99.2%)	6 (0.6%)	2 (0.2%)
Vomiting	918 (98.7)	8 (0.9)	4 (0.4)	981 (98.3%)	15 (1.5%)	2 (0.2%)

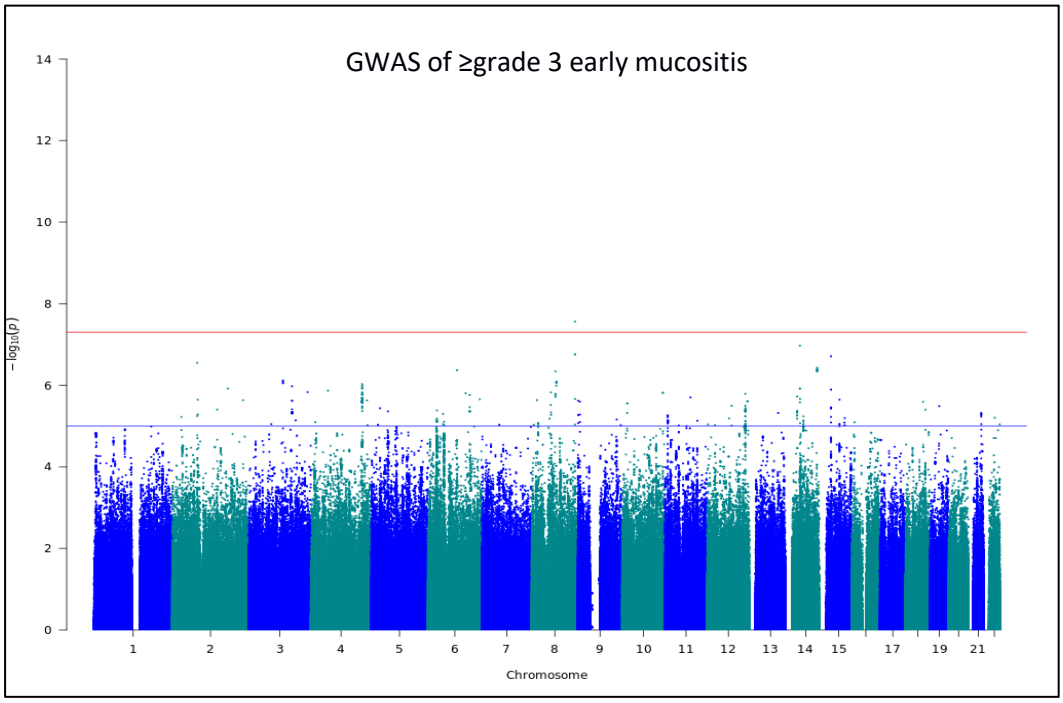
The frequency of toxicity cases during the first two cycles for each toxicity variable according to CTCAE grading is shown. NA- not available.

rs181841887 (associated with early ≥grade 3 HFS toxicity in QUASAR2, OR=6.41, 95% CI=3.33-12.22, $P=2.6x10^{-8}$) could not be validated in COIN because it was neither directly genotyped

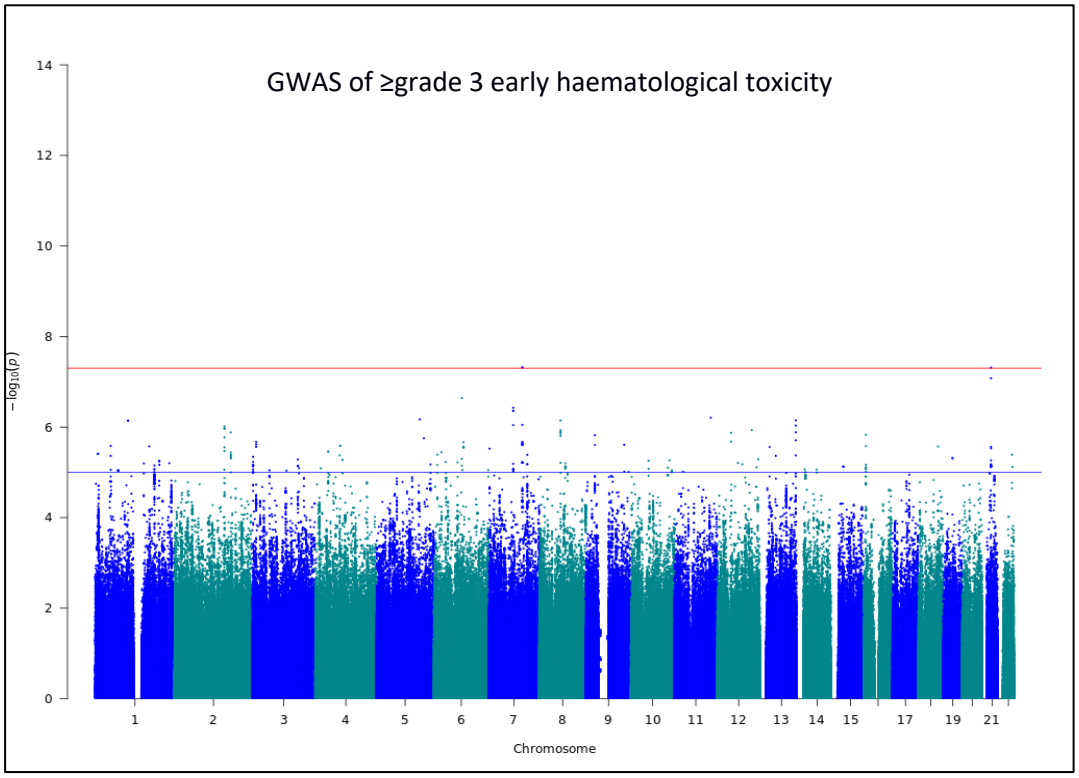
nor imputed with sufficient accuracy. Two SNPs, rs179893 and rs200507264, were associated with early ≥grade 3 haematological toxicity in the COIN dataset (both OR=23.83, 95% CI=7.63-74.41, P=4.82x10^{-8}) but could not be validated in QUASAR2.

Meta-analyses combining the QUASAR2 and COIN datasets was performed to identify genetic regions associated with the risk of any capecitabine-induced early-onset ≥grade 3 toxicity (global) and individual toxicities, including ≥grade 3 mucositis, vomiting, and haematological toxicities. Although grade 2 toxicity events would normally also lead to treatment interruptions in clinical practice, first occurrence of these events are not usually dose-limiting unless they are recurrent. Hence, ≥grade 3 events has been selected as "cases" in the association analyses conducted as these normally leads to dose reductions and discontinuation in practice. **Figure 3.1 (a-d)** shows the Manhattan plots generated from the meta-analysis of four early toxicity phenotypes with GWAS significant hits. The Manhattan plot of early global toxicity shows a single locus reaching genome-wide significance. Multiple SNPs in linkage disequilibrium (LD) with one another support an association between this locus and global toxicity.

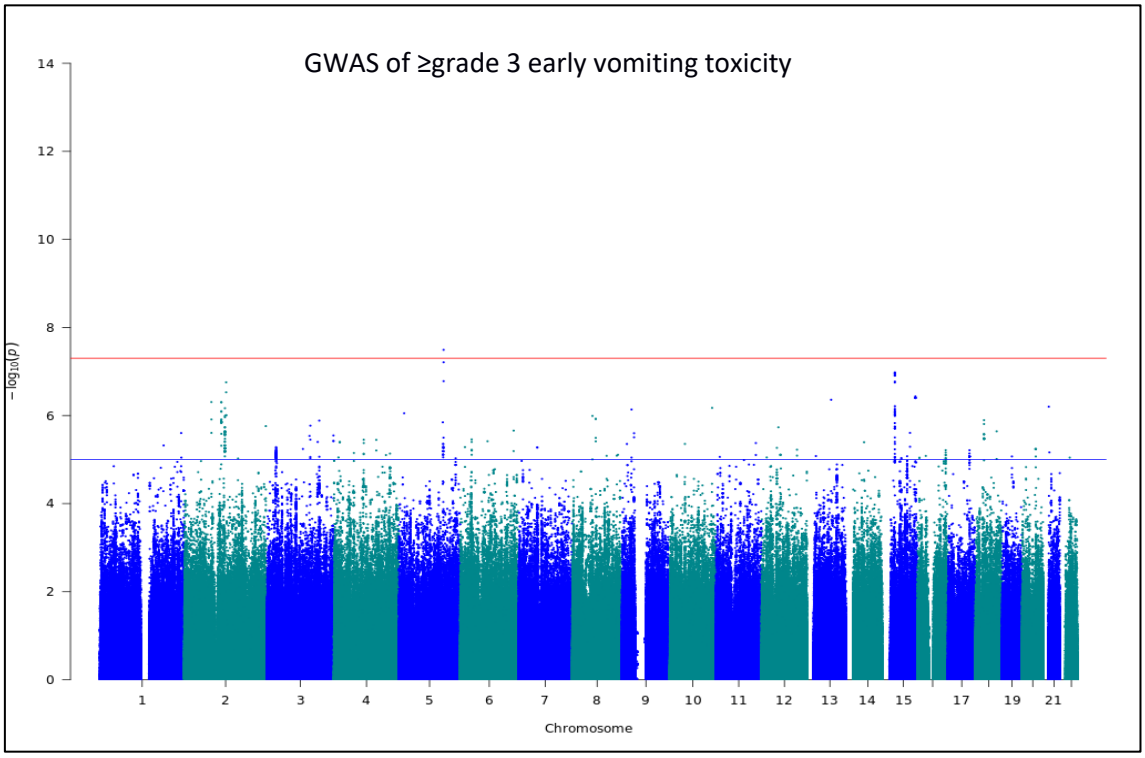
a)


GWAS of ≥grade 3 early mucositis
14
12
10
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
19
21
Chromosome

b)


GWAS of ≥grade 3 early haematological toxicity
14
12
10
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
19
21
Chromosome

c)



d)

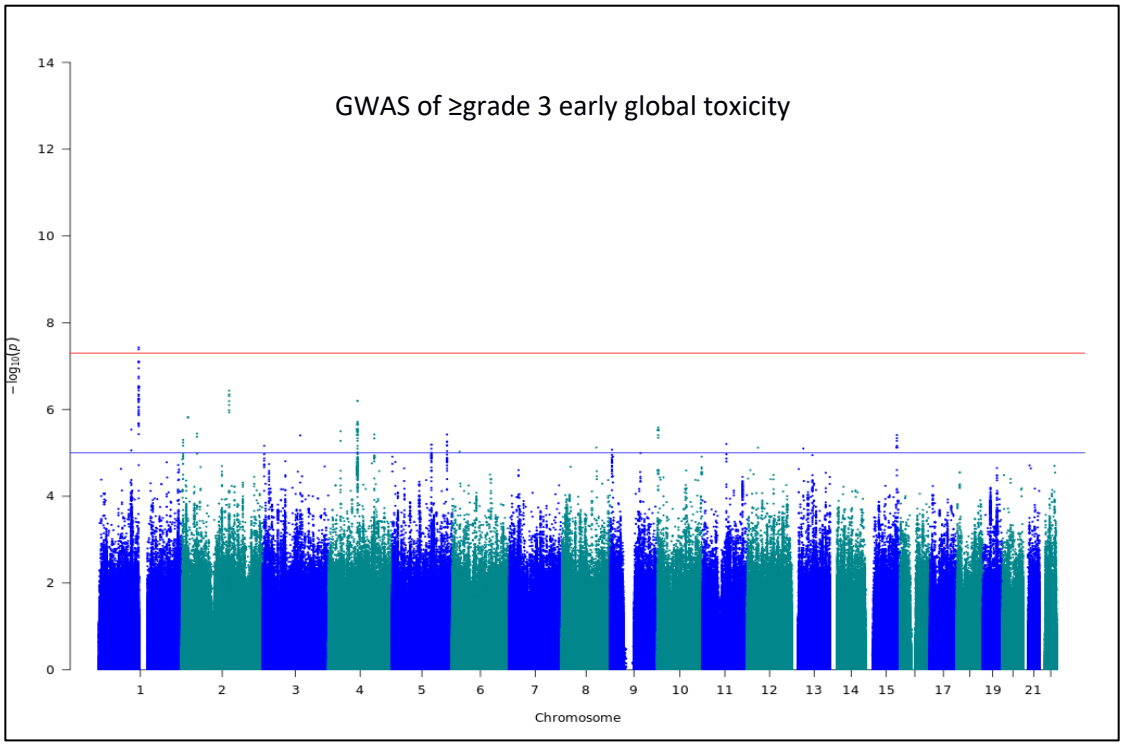


Figure 3.1 Manhattan plots showing results from the GWAS meta-analysis of ≥grade 3 capecitabine toxicity in QUASAR2 and COIN within the first two treatment cycles

GWAS outputs for the meta-analysis of early ≥grade 3 a) Mucositis, b) Haematological, c) Vomiting and d) Global (any) toxicity is shown. For each plot, the x-axis represents the SNPs from different chromosomes and the y-axis represents the significance level of association at -log10 of the P value of each SNP with the presence of any early ≥grade 3 capecitabine toxicity. The red and blue lines represent the P value threshold of $<5x10^{-8}$ (GWAS significance level) and $<5x10^{-5}$. Abbreviations: GWAS- genome-wide association study

Summary statistics of the top independent SNPs identified from the GWAS meta-analysis of both QUASAR2 trial arms and the COIN study are shown in **Table 3.9**. Four SNPs with MAFs of <10% reached GWAS significance at $P<5x10^{-8}$ in their association with any (global) or individual ≥grade 3 early capecitabine toxicity. These SNPs are not previously known to be associated with capecitabine toxicity. There were no GWAS significant hits for diarrhoea and HFS from the meta-analysis of both trial cohorts.

Table 3.9 Top SNPs identified from the meta-analysis of QUASAR2 and COIN datasets of early ≥grade 3 capecitabine toxicity under the additive genetic model (total N=1928)

Toxicity	Chr	Pos	rsIDs	Risk allele	Freq	OR	Lower CI	Upper CI	P value	Consequences	Mapped gene/nearest genes
Mucositis	8	137378694	rs117241924	A	0.02	28.82	8.81	94.30	2.74E-08	intergenic	KHDRBS3(dist=718842),LINC02055(dist=164347)
Haem	21	27498400	rs140805091	G	0.02	14.21	5.48	36.86	4.85E-08	intronic	APP
Vomiting	5	131697770	rs146950191	A	0.02	21.34	7.21	63.16	3.22E-08	ncRNA intronic	AC034220.3
Global	1	119858910	rs2477888	A	0.08	2.16	1.64	2.83	3.71E-08	ncRNA intronic	RP11-418J17.3

Conditional and joint ('COJO') analysis was used to identify top independent SNPs associated with each toxicity variable that reached a P-value threshold of $<5x10^{-8}$ after filtering for MAF >1%, info score > 0.8, HWE $1x10^{-6}$, missingness 3% and P heterogeneity >0.05. The risk allele refers to the allele associated with toxicity risk. All SNPs are present in COIN and at least one arm of QUASAR2 unless indicated otherwise. Abbreviations: Chr -chromosome, CI- confidence interval, Freq -frequency, Haem-haematological, ncRNA- non-coding RNA, OR -odds ratio, Pos- genomic position (build GRCh37)

An expression quantitative trait loci (eQTL) is a locus that explains a proportion of the genetic variance of a gene expression phenotype[174]. To assess the implications of the top toxicity-associated SNPS on changes in gene expression in specific tissues, eQTL mapping was performed using FUMA GWAS v1.5.1 to evaluate the association between the SNPs with gene expression levels typically measured in hundreds of individuals[174].

eQTL data of the top SNPs identified for each toxicity phenotype is shown in **Table 3.10**. rs140805091(A>G) associated early haematological toxicity is a significant eQTL for *APP* expression ($P=1.8x10^{-5}$) in the whole blood. rs146950191 associated with vomiting toxicity is a significant eQTL for *SLC22A5* and *AC034220.3* where its risk-increasing 'A' allele is associated with increased expression of *AC034220.3* in the brain cortex ($P_{corrected} = 2.2x10^{-23}$) and skin (lower leg) ($P_{corrected} =1.1x10^{-15}$). *SLC22A5* encodes an organic cation and carnitine transporter and *AC034220.3* has four transcripts (splice variants)[207]. rs2477888 (C>A) associated with any ≥grade 3 (global) toxicity during the first two treatment cycles is a significant eQTL of *RP11-418J17.1* and increases *RP11-418J17.1* expression (lowest $P_{corrected} = 1x10^{-16}$) in brain tissue. The SNP is also a significant eQTL of *WARS2* and decreases the expression of *WARS2* in B cells ($P_{corrected}=4.9x10^{-3}$) and monocytes (lowest $P_{corrected}=6.1x10^{-5}$). No significant eQTLs were detected from the GWAS meta-analysis of early mucositis.

Table 3.10 Expression quantitative trait loci (eQTL) data from top SNPs from the GWAS meta-analysis of early capecitabine toxicity

Toxicity type	SNP	Risk allele	Gene Symbol	Aligned Direction	Tissue	Value	FDR	eQTL database
Haem	rs140805091	G	APP	+	Neutrophil	2.62E-07	1.30E-03	eQTLcatalogue (BLUEPRINT)
		G	APP	-	Whole blood	1.77E-05	2.61E-11	GTEx/v8
Vomiting	rs146950191	A	SLC22A5	+	Multiple tissues	7.43E-63	0	eQTLGen
		A	AC034220.3	+	Multiple tissues	1.89E-14	0	eQTLGen
		A	SLC22A5	+	Artery- Tibial	2.48E-04	6.23E-22	GTEx/v8
		A	AC034220.3	+	Brain- Cortex	2.20E-05	2.18E-23	GTEx/v8
		A	SLC22A5	+	Nerve- Tibial	2.85E-04	1.35E-08	GTEx/v8
		A	AC034220.3	+	Skin- Sun-exposed lower leg	1.50E-04	1.07E-15	GTEx/v8
Global	rs2477888	A	WARS2	-	B-cell- CD19	9.79E-07	4.89E-03	eQTLcatalogue (FAIRFAX 2012)
		A	WARS2	-	Monocytes- IFN24	1.22E-06	6.10E-03	eQTLcatalogue (FAIRFAX 2014)
		A	WARS2	-	Monocytes- LPS24	2.28E-06	1.14E-02	eQTLcatalogue (FAIRFAX 2014)
		A	WARS2	-	Monocytes- Naive	1.23E-08	6.14E-05	eQTLcatalogue (FAIRFAX 2014)
		A	WARS2	+	Multiples tissues	4.11E-55	0	eQTLGen
		A	RP11-418J17.1	+	Multiples tissues	4.29E-07	1.36E-03	eQTLGen
		A	RP11-418J17.1	+	Brain- Frontal Cortex	1.01E-05	1.02E-16	GTEx/v8
		A	RP11-418J17.1	+	Brain- Substantia nigra	8.34E-06	1.18E-07	GTEx/v8

Unique pairs of SNP-gene-tissue are shown for the top independent SNPs associated with early-onset capecitabine toxicity. eQTL mapping was performed using FUMA GWAS based on the data sources listed[84]. Aligned direction is the direction of effect on gene expression, where "+" (risk allele increases gene expression) or "-" (risk allele decreases gene expression). FDR- False discovery rate of eQTLs after multiple testing. eQTL database- data source of eQTLs used: eQTLGen[175], eQTLcatalogue (FAIRFAX 2012 dataset includes B cells based on 282 samples and donors and FAIRFAX 2014 includes monocytes based on 1372 samples and 424 donors)[208][209], GTEx/v8[210]

To determine whether the top toxicity-associated genetic markers involve specific biological or cellular function pathways, gene-set analysis was performed using curated gene sets and gene ontology terms from MSigDB using MAGMA [172,211]. The gene sets consist of groups of genes sharing certain biological or functional characteristics. **Table 3.11** shows the top 10 gene sets from the GWAS of early ≥grade 3 global toxicity with the most significant gene sets including negative regulation of B cell proliferation (OR=2.6, $P=1.4x10^{-5}$, N=15 genes), transposition biological processes (OR=1.8, $P=8.6x10^{-5}$, N=28 genes) and amino acid binding molecular function (OR=1.4, $P=1.9x10^{-4}$, N=81 genes).

Table 3.11 Top 10 gene sets from MAGMA gene-set analysis from GWAS meta-analysis of early ≥grade 3 global capecitabine toxicity

Gene Set	N genes	Beta	SE	P	P bon
GO BP: negative regulation of B cell proliferation	15	0.957	0.229	1.43E-05	0.221
GO BP: transposition	28	0.587	0.156	8.56E-05	1
GO MF: modified amino acid binding	81	0.332	0.094	1.98E-04	1
Curated gene sets: Wong endometrium cancer	23	0.546	0.155	2.04E-04	1
GO MF: cytokine binding	119	0.280	0.080	2.45E-04	1
GO MF: LIM domain binding	7	0.898	0.268	3.98E-04	1
Curated gene sets: reactome synthesis of 12 eicosatetraenoic acid derivatives	7	0.904	0.270	4.05E-04	1
GO MF: oligopeptide binding	12	0.735	0.225	5.51E-04	1
GO CC: SCF ubiquitin ligase complex	62	0.331	0.102	5.73E-04	1
GO BP: regulation of actin nucleation	27	0.457	0.141	5.75E-04	1

Using FUMA, the curated gene sets (C2.all) and gene ontology (GO) terms (C5. BP- Biological Processes, CC- Cellular Component and MF: Molecular Functions ontologies) are analysed [84,211]. Gene sets from MsigDB v6.2 are used[172]. Abbreviations: Beta- regression coefficient of the gene set, N genes- number of genes in the data that are in the gene set, P bon- P value after Bonferroni correction for multiple testing for all tested gene sets, SCF- SKP1-Cullin/Cdc53-F-box protein, SE- standard error of the regression coefficient

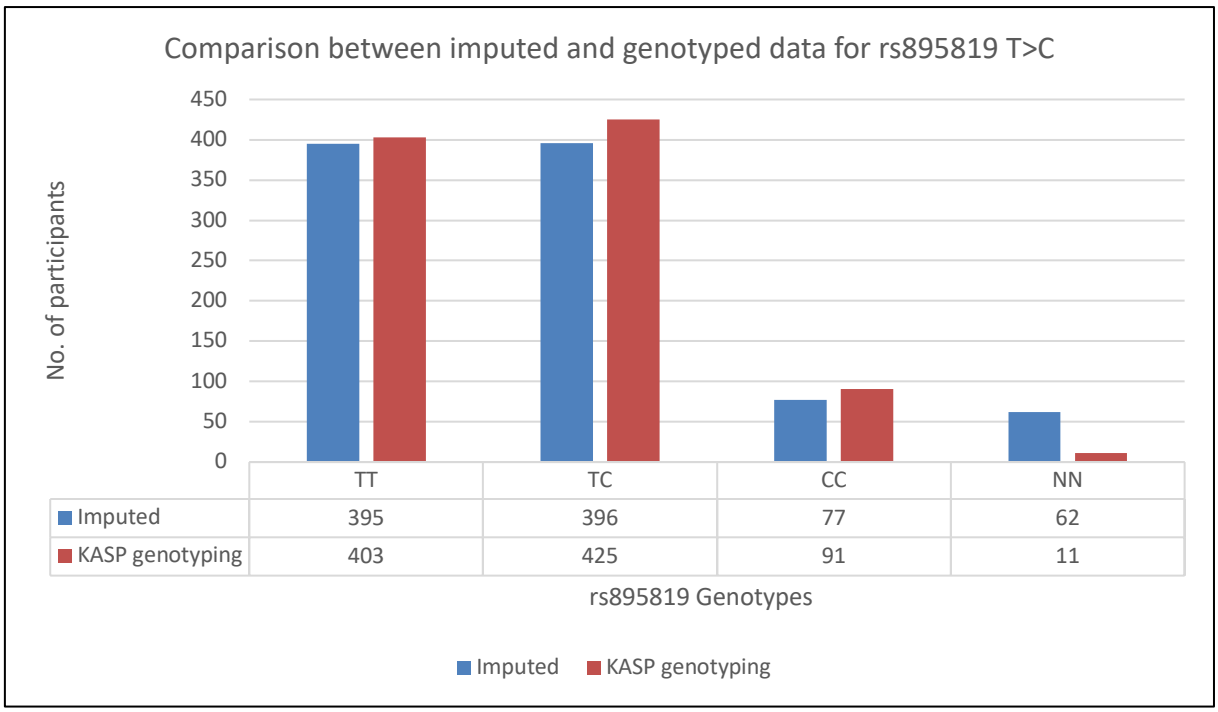
3.2.8 Association between *MIR27A* SNPs with early-onset fluoropyrimidine toxicity

3.2.8.1 KASP genotyping for both MIR27A SNPs in QUASAR 2

MIR27A SNPs rs895819 and rs11671784 were previously investigated as markers of early-onset fluoropyrimidine induced toxicity[115]. Given that both SNPs were imputed in QUASAR2 with INFO scores of 0.96 and 0.62 and missing data rates of 7% and 5% respectively, I decided to genotype these SNPs using KASP (LGC Genomics) genotyping assays. Sanger sequencing was used to confirm a representative sample of each observed genotype.

To compare the results of imputation and direct genotyping, I hard-coded the imputed data calling a genotype only if the probability of that genotype was ≥0.9. If all possible genotypes had a probability <0.9, the genotype was coded as missing. **Figure 3.2 a-b** shows the imputation-derived and KASP assay-derived genotypes for both *MIR27A* variants**.** 77 and 91 participants are homozygous carriers for the risk allele (C/C) of rs895819 from imputation-derived and KASP genotyping respectively. Two participants were carriers of both risk alleles (A/A) of the rarer SNP rs11671784 from KASP genotyping, while there were no homozygous risk allele carriers from imputed data. Missing data rates were 1% for both SNPs after genotyping.

a)



b)

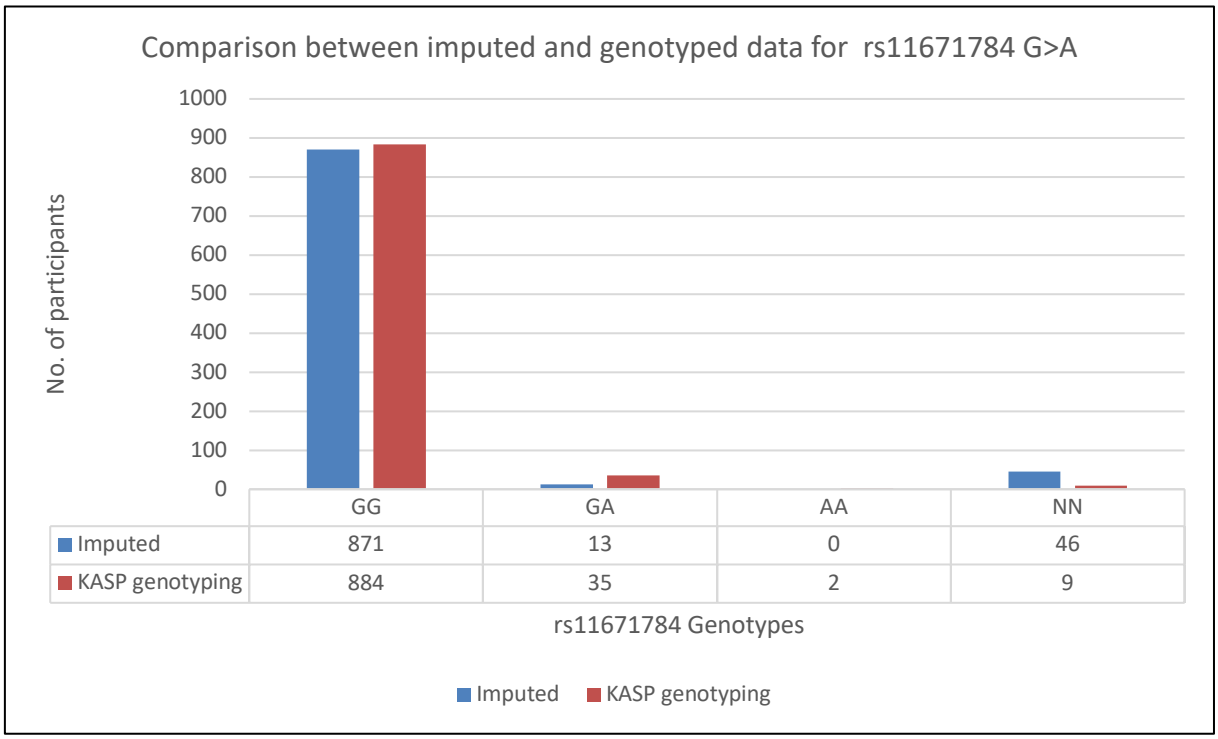


Figure 3.2 Genotypic profile of two *MIR27A* SNPs in the QUASAR2 trial cohort according to imputation or KASP genotyping derived genotypes

The number of homozygous and heterozygous carriers for the *MIR27A* common variant rs895819 (risk allele "C") and rare variant rs11671784 (risk allele "A") are shown. Participants carrying the rs895819 T/T and rs11671784 G/G genotypes are homozygous carriers for the non-risk alleles, those with rs895819 T/C and rs11671784 G/A genotypes are heterozygous carriers and those with rs895819 C/C and rs11671784 A/A genotypes are homozygous carriers for the risk alleles. Samples that have undetermined genotyping results or are not available for genotyping (N/N) are classed as missing.

3.2.8.2 Association between common MIR27A SNP rs895819 and early and all cycle ≥grade 3 toxicity in QUASAR2

Association analysis of *MIR27A* rs895819 T>C variant with each listed toxicity type was conducted as described in Methods Chapter 2.3.2. Both arms of the QUASAR2 trial (Arm A and Arm B) were initially analysed separately and then meta-analysed. Results of the meta-analyses of both trial arms for cycles 1 and 2 using imputed or KASP-directed genotyping data are shown in **Table 3.12(a)**. As can be seen, similar results were observed in the imputed and genotyped data. No significant associations with any of the toxicity phenotypes were observed.

To confirm that imputation artefacts were not driving the association between rs895819T>C and diarrhoea or GI toxicities across all treatment cycles, I repeated the analysis using the genotyped data. As shown in **Table 3.12(b)**, similar results were observed with the genotyped data and the associations with diarrhoea or GI toxicities remain statistically significant. The suggestive association with haematological toxicity has disappeared.

Table 3.12 Summary statistics of the association between rs895819 T>C variant and ≥grade 3 capecitabine-induced toxicities using both imputed and KASP genotyping data

a)

Early-onset	Imputed			KASP genotyping		
	OR	95% CI	P value	OR	95% CI	P value
Diarrhoea	1.33	0.88-1.98	0.167	1.37	0.92-2.02	0.118
Haem	0.71	0.26-1.92	0.506	0.76	0.28-2.05	0.584
HFS	1.06	0.73-1.54	0.760	1.14	0.79-1.63	0.485
Mucositis	1.15	0.43-3.07	0.779	1.4	0.55-3.60	0.484
Vomiting	1.39	0.47-4.12	0.552	1.25	0.44-3.55	0.681
GI	1.26	0.86-1.85	0.243	1.29	0.88-1.88	0.186
Global	1.17	0.88-1.56	0.272	1.21	0.92-1.60	0.175

b)

All cycles	Imputed			KASP genotyping		
	OR	95% CI	P value	OR	95% CI	P value
Diarrhoea	1.55	1.11-2.15	0.009	1.58	1.15.2.18	0.005
Haem	0.53	0.23-1.24	0.145	0.75	0.35-1.64	0.477
HFS	0.88	0.68-1.13	0.311	0.98	0.77-1.25	0.888
Mucositis	1.15	0.48-2.76	0.751	1.32	0.58-3.02	0.514
Vomiting	0.92	0.40-2.11	0.852	0.86	0.39-1.93	0.723
GI	1.48	1.07-2.03	0.016	1.50	1.10-2.04	0.010
Global	1.01	0.81-1.26	0.906	1.11	0.90-1.38	0.337

Association testing results for each toxicity variable are shown based on the additive genetic model across a) treatment cycles 1 and 2 and b) all treatment cycles. Gastrointestinal toxicity is defined as the presence of diarrhoea and/or mucositis. Abbreviations: CI- confidence interval, GI- gastrointestinal, Haem- haematological, HFS- hand-foot-syndrome, OR- odds ratio

3.2.8.3 Association between rare MIR27A SNP rs11671784 with early cycle and all cycle ≥grade 3 toxicity in QUASAR2

With the genotype data available, it was now possible to assess the rare *MIR27A* variant (MAF 2%) for associations with any-cycle and early-onset (cycles 1 and 2) ≥grade 3 toxicity in QUASAR2 with more confidence. Meta-analyses results of both trial arms using the KASP genotyping data of cycles 1 and 2 toxicity are shown in **Table 3.13**. As the number of patients with ≥grade 3 toxicity cases who carry the toxicity risk allele of rs11671784 is low, only the results for the global toxicity variable analysis are shown. No significant associations were observed.

Table 3.13 Summary statistics of the association between rs11671784 G>A variant and any (global) ≥grade 3 capecitabine toxicity using KASP genotyping data

	Capecitabine-induced global toxicity		
	OR	95% CI	P value
Early-onset	0.97	0.36-2.60	0.952
Any-cycle	1.84	0.90-3.77	0.095

Association testing results are shown based on the additive genetic model for treatment cycles 1 and 2 (early-onset) and all treatment cycles. Abbreviations: CI- confidence interval, OR- odds ratio

3.2.8.4 Meta-analysis investigating the association between MIR27A SNPS and early ≥grade 3 toxicity in QUASAR2 and Dutch cohorts

Initial findings from QUASAR2 showed suggestive associations between *MIR27A* variants rs895819 with early-onset ≥grade 3 capecitabine toxicity, especially GI toxicity, but did not reach nominal significance. This was particularly the case for the rare *MIR27A* variant rs11671784 due to its low MAF. Hence, meta-analyses were conducted combining datasets from the QUASAR2 and a Dutch cohort from collaborator Didier Meulendijks to increase the sample size and power of the analysis to assess if there was any significant association between both *MIR27A* variants and early ≥grade 3 fluoropyrimidine toxicity. Both studies were large datasets that had collected similar toxicity data from patients treated with fluoropyrimidine chemotherapy. Binary variables were generated for each toxicity variable with ≥grade 3 events as cases and grade 0-2 events as controls.

The Dutch cohort involved 1631 patients enrolled to a prospective, multicenter study in Netherlands with pan-cancer diagnosis intending to receive fluoropyrimidine-based chemotherapy either as monotherapy or combination with other chemotherapy or radiotherapy[163]. Germline DNA of the patients were obtained prospectively and genotyped for *DPYD**2A before starting treatment[163]. 18 patients identified to be carriers of the *DPYD**2A variant was excluded from our collaborator's analysis as they were treated with a reduced dose of fluoropyrimidine[115]. Early toxicity data (first two treatment cycles) and genotyping data for the two *MIR27A* SNPs rs895819 and rs11671784 were available for analysis from 1592 study participants ('All Participants' cohort). 1427 out of the 1592 study participants were treated with capecitabine ('Capecitabine Only' cohort), while 165 participants were treated with intravenous (IV) 5-FU based therapies. The number of participants with ≥grade 3 toxicity events in cycles 1 and 2 is shown in **Table 3.14**.

Table 3.14 Number of cases in the Dutch cohort with ≥grade 3 toxicities in treatment cycles 1 and 2

Toxicity type	No. of ≥ grade 3 events in the 'All Participants' cohort (n=1592)	No. of ≥ grade 3 events in the 'Capecitabine Only' cohort (n=1427)
Diarrhoea	100	94
Haem	95	79
HFS	44	44
Mucositis	22	18
GI	115	105
Global	230	205

The toxicity cases in participants treated with either intravenous 5-FU or capecitabine based therapies ('All Participants') cohort and those who received capecitabine-based therapies ('Capecitabine Only') cohort is shown. Abbreviations: 5-FU- fluorouracil, GI- gastrointestinal, Haem- haematological, HFS- hand-foot-syndrome, IV- intravenous

Table 3.15 shows the results of the meta-analyses for common *MIR27A* variant rs895819 using the additive genetic model. There was a significant association between the rs895819T>C variant and increased risk of diarrhoea and global toxicity during the first two treatment cycles in both the "All Participants" and "Capecitabine Only" study cohorts. The association with GI toxicity reached nominal significance in the capecitabine-only analysis.

Table 3.15 Meta-analysis of rs895819 with early ≥grade 3 toxicity from QUASAR2 and the Dutch study cohort (additive model)

Early Toxicity	All Participants (N=2522)				Capecitabine Only (N=2357)			
	No. of grade 3-4 cases	OR	95% CI	P value	No. of grade 3-4 cases	OR	95% CI	P value
Diarrhoea	164	1.29	1.01-1.64	0.036	158	1.35	1.06-1.72	0.016
Haem	107	0.96	0.71-1.30	0.799	91	0.94	0.68-1.31	0.720
HFS	121	1.19	0.90-1.58	0.211	121	1.2	0.90-1.58	0.213
Mucositis	32	1.06	0.62-1.81	0.825	28	1.23	0.70-2.15	0.472
GI	185	1.24	0.99-1.56	0.058	175	1.33	1.05-1.68	0.017
Global	370	1.19	1.00-1.41	0.044	345	1.23	1.03-1.46	0.021

Association testing to assess the association between rs895819 and early ≥grade 3 toxicity (treatment cycles 1 and 2) was performed using the additive genetic model. The 'All participants' cohort received either IV 5-FU or capecitabine based therapies and the 'Capecitabine only' cohort received capecitabine based therapies only. Abbreviations: CI- confidence interval, GI- gastrointestinal, Haem- haematological, HFS- hand-foot-syndrome, OR- odds ratio

Given that Meulendijks et al had previously assess the association of the *MIR27A* variants with early toxicity under the dominant genetic model[115], I repeated the meta-analyses to validate the effects of the variants under the dominant model and also test their effects in the other commonly used recessive model for GWAS in our meta-analysis. The dominant model did not show any associations reaching nominal significance. Results of the meta-analyses under the recessive model is shown in **Table 3.16**, of which the significance level of the rs895819 toxicity-associated allele was the strongest. The GI analysis is also statistically significant under the recessive model. The variant continues to have a protective effect on haematological toxicity under the additive and recessive model across both cohorts, although this does not reach nominal significance.

Table 3.16 Meta-analysis of rs895819 with early ≥grade 3 toxicity from QUASAR2 and the Dutch study cohort (recessive model)

Early Toxicity	All Participants (N=2522)				Capecitabine only (N=2357)			
	OR	Lower CI	Upper CI	P value	OR	Lower CI	Upper CI	P value
Diarrhoea	1.68	1.19	2.36	0.003	1.79	1.26	2.55	0.001
Haem	0.93	0.62	1.38	0.703	0.91	0.59	1.39	0.654
GI	1.55	1.13	2.13	0.007	1.71	1.23	2.39	0.002
Global	1.36	1.08	1.71	0.008	1.43	1.13	1.82	0.003

Association testing to assess the association between rs895819 and early ≥grade 3 toxicity (treatment cycles 1 and 2) was performed using the recessive genetic model. The 'All Participants' cohort received either IV 5-FU or capecitabine-based therapies and the 'Capecitabine Only' cohort received capecitabine based therapies only. Abbreviations: CI- confidence interval, GI- gastrointestinal, Haem- haematological, HFS- hand-foot-syndrome, OR- odds ratio

3.2.8.5 Meta-analysis of the association between MIR27A SNP rs11671784 and early ≥grade 3 toxicity in QUASAR2 and Dutch study cohorts

Meta-analyses were also conducted to assess the association between the rare *MIR27A* SNP rs11671784 G>A and early toxicity in the UK and Dutch cohorts. Results are shown in **Table 3.17**. No significant associations were observed.

Table 3.17 Meta-analysis of rs11671784 with early ≥grade 3 toxicity from QUASAR2 and the Dutch study cohort

Early Toxicity	All Participants (N=2522)				Capecitabine only (N=2357)			
	No. of grade 3-4 cases	OR	95% CI	P value	No. of grade 3-4 cases	OR	95% CI	P value
Diarrhoea	164	0.75	0.23-2.45	0.636	158	0.78	0.24-2.55	0.682
Haem	107	1.50	0.57-3.90	0.412	91	1.25	0.36-4.33	0.729
HFS	121	1.28	0.54-3.06	0.572	121	1.29	0.54-3.08	0.561
Mucositis	32	1.12	0.15-8.47	0.914	28	1.38	0.18-10.61	0.754
GI	185	0.87	0.31-2.46	0.799	175	0.95	0.33-2.68	0.918
Global	370	0.98	0.55-1.76	0.950	345	0.89	0.35-2.27	0.715

Association testing to assess the association between rs11671784 and early ≥grade 3 toxicity (treatment cycles 1 and 2) was performed using the additive genetic model. The 'All Participants' cohort received either IV-5FU or capecitabine-based therapies and the 'Capecitabine Only' cohort received capecitabine based therapies only. Abbreviations: CI- confidence interval, GI- gastrointestinal, Haem- haematological, HFS- hand-foot-syndrome, OR- odds ratio

3.2.8.6 Stratified analyses between MIR27A rs895819 and early ≥grade 3 toxicity according to DPYD carrier status

Given the significant association between rs895819A>G with GI and global toxicity in the previous meta-analyses, the analyses were repeated in both QUASAR2 and Dutch cohorts stratified according to *DPYD* risk variants carrier status. **Table 3.18** shows the summary statistics results for the two groups under the additive genetic model. rs895819 polymorphism was significantly associated with an increased risk of early GI and global toxicity in *DPYD* wild type patients and not in the smaller subset with a *DPYD* risk allele. It was not possible to analyse the dominant model in the DPYD risk allele carriers due to the low sample size.

Table 3.18 Meta-analysis of rs895819 with early ≥grade 3 diarrhoea, GI and global toxicity in QUASAR2 and Dutch cohorts according to *DPYD* carrier status

Toxicity	DPYD risk allele carriers (n=130)				DPYD wild type (n=2306)			
	OR	Lower CI	Upper CI	P-value	OR	Lower CI	Upper CI	P- value
Diarrhoea	1.24	0.43	3.57	0.68	1.33	1.04	1.71	0.02
GI	0.88	0.35	2.20	0.79	1.30	1.03	1.65	0.03
Global	0.88	0.40	1.92	0.75	1.23	1.03	1.47	0.02

Association analysis of the risk of early ≥grade 3 toxicity in carriers and non-carriers of any CPIC *DPYD* risk variants under the additive model are shown here. Abbreviations: CI- 95% confidence interval GI- gastrointestinal, OR- odds ratio

3.2.8.7 Sensitivity analyses of MIR27A rs895819 in predicting early toxicity

The ability of the common *MIR27A* SNP rs895819 to detect patients at risk of ≥grade 3 early toxicity was analysed to assess its clinical utility. As shown in **Table 3.19**, the presence of rs895819 variant alone had a sensitivity of 86%, 79% and 81% in predicting early-onset ≥grade 3 diarrhoea, GI and any (global) toxicity respectively. All three analyses achieved a specificity of 30% in predicting these events. When combined with the presence of any of the *DPYD* risk alleles, a high specificity of up to 96% but low sensitivity of ~10% was observed in detecting these toxicity events.

Table 3.19 Predictive ability of MIR27A rs895819 and CPIC DPYD risk variants in determining the risk of early ≥grade 3 capecitabine toxicity

Toxicity	Variants	Sensitivity	Specificity	PPV	NPV
Diarrhoea	*DPYD* alone	10.61%	94.96%	10.77%	94.88%
	DPYD + rs895819	9.85%	96.13%	12.75%	94.89%
	rs895819 alone	85.61%	29.77%	6.54%	97.30%
GI	*DPYD* alone	11.54%	95.16%	16.15%	93.01%
	DPYD + rs895819	10.44%	96.31%	18.63%	93.01%
	rs895819 alone	79.12%	29.59%	8.33%	94.60%
Global	*DPYD* alone	10.47%	95.32%	22.31%	89.23%
	DPYD + rs895819	10.11%	96.57%	27.45%	89.32%
	rs895819 alone	81.23%	30.24%	13.01%	92.61%

The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) in predicting early diarrhoea, GI and any (global) ≥grade 3 toxicity in either *DPYD* or rs895819 risk variant carriers alone or DPYD risk variant carriers also carry the rs895819 risk allele (DPYD + rs895819T>C). GI toxicity refers presence of diarrhoea and/or mucositis.

3.2.9 Polygenic risk profiling

To assess the predictive abilities of the top independent SNPs in identifying patients at risk of capecitabine-induced ≥grade 3 toxicities, polygenic risk scores (PRS) were generated based on top SNPs identified from the QUASAR2 study. These SNPs were detected based on the "expected" method approach to handle genotype uncertainty and reached a suggestive GWAS significance P-value threshold of $5x10^{-7}$. The sensitivity and specificity of these SNPs were measured using the area under the receiver operating characteristic curve (AUC-ROC). As we did not have access to genotyping data for an independent dataset, the dataset used to test the predictive value of the PRS was the same dataset used to generate the PRS (both from the QUASAR2 trial), hence there will be a substantial risk of overfitting that needs to be considered. Results for AUC and sensitivity analyses for the top SNPs identified for each toxicity variable in both early-onset and all treatment cycles are shown in **Figure 3.3-3.5 (A-D).** The top SNPs associated with mucositis toxicity were found to have the highest AUC in predicting ≥grade 3 mucositis events in the first two treatment cycles (AUC 98.7%) and across all treatment cycles (AUC 98%) **(Figure 3.3 A-B)**. When all top 26 SNPs associated with both early or any-cycle ≥grade 3 toxicity of any type were considered, an AUC of 72.8% with a sensitivity of 64.8% and specificity of 69.5% was achieved for predicting any early onset ≥grade 3 toxicity (global) at a PRS threshold of -0.005. The same set of SNPs achieved an AUC of 63.9% with a sensitivity of 47.7% and specificity of 75.9% for predicting any ≥grade 3 toxicity across all cycles at a PRS threshold of 0.004 **(Figure 3.5 C-D)**.

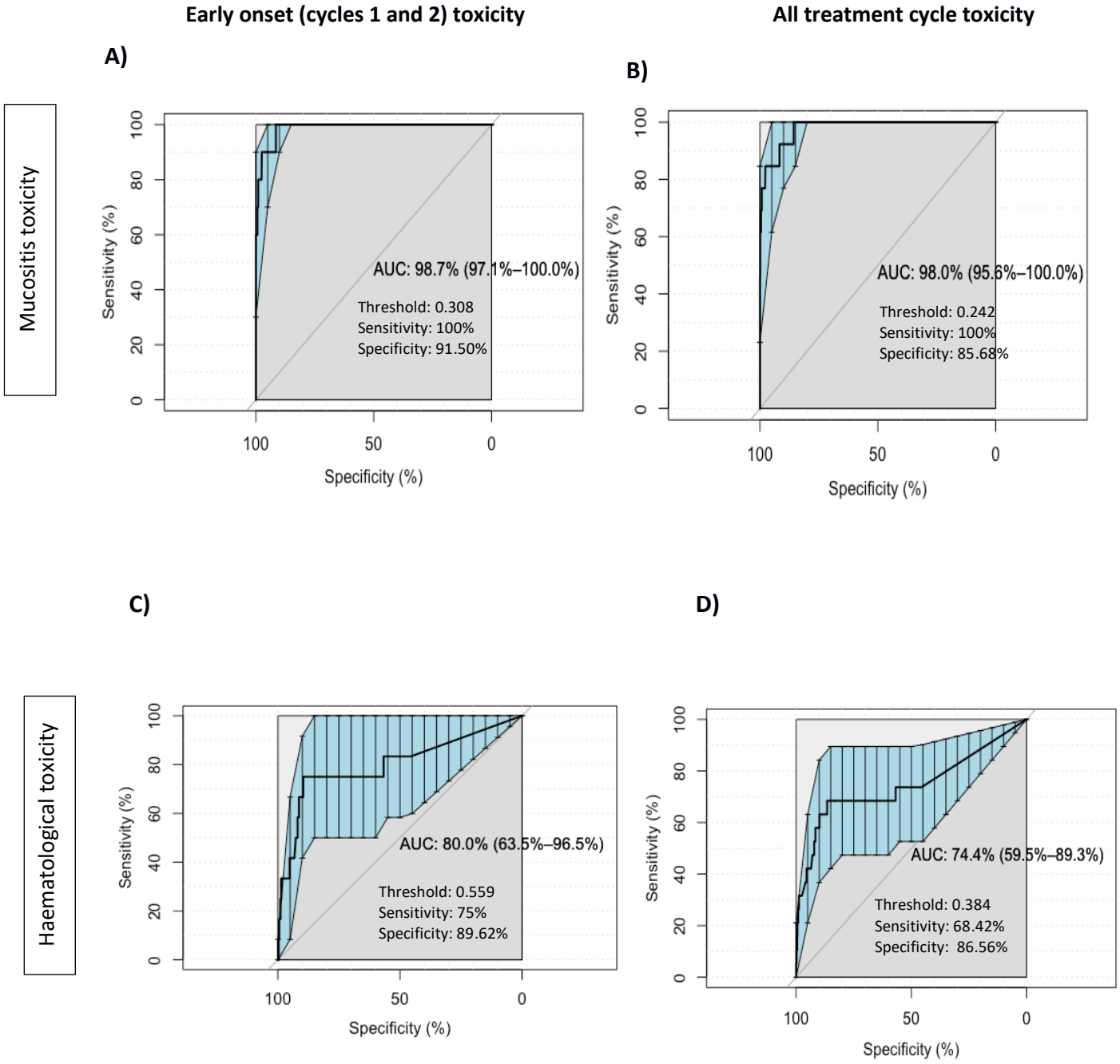


Figure 3.3 AUC-ROC of the top independent SNPs associated with mucositis and haematological toxicity

ROC analysis results of top 5 SNPs associated with mucositis (rs76211390, rs79272399, rs77911328, rs145225093 and rs117241924) in predicting **(A)** early onset ≥grade 3 mucositis toxicity and **(B)** any cycles ≥grade 3 mucositis toxicity at the respective threshold indicated. ROC analysis results for the top 2 SNPs associated with haematological toxicity (rs75217739 and rs139788123) in predicting **(C)** early onset ≥grade 3 haematological toxicity and **(D**) any cycle ≥grade 3 haematological toxicity at the respective threshold indicated. The error bars represent the confidence intervals of the curve. Abbreviations: AUC- area under the curve, ROC- receiving operating characteristic curve

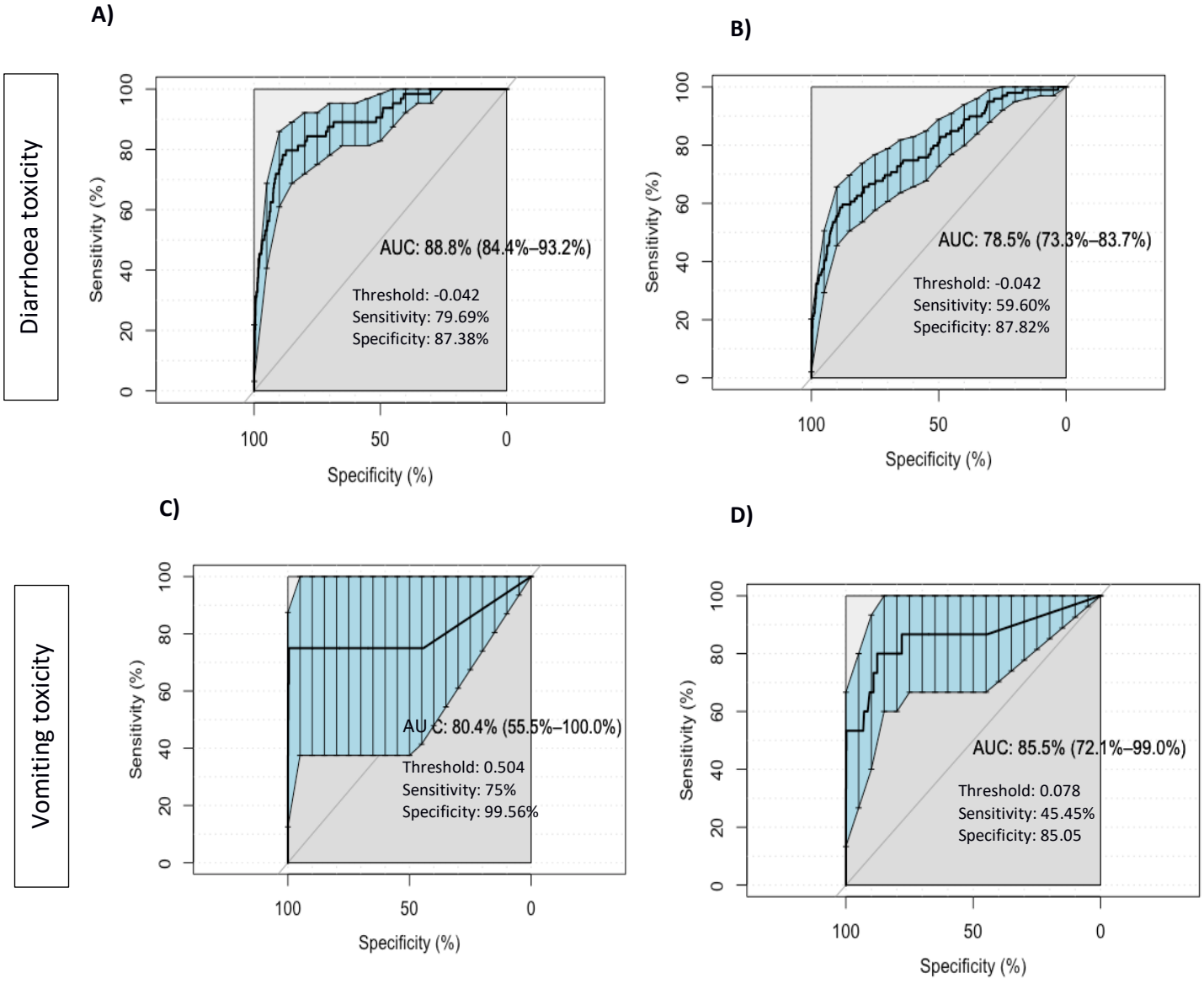


Figure 3.4 AUC-ROC of the top independent SNPS associated with diarrhoea and vomiting toxicity
ROC analysis results for the top 10 SNPs associated with diarrhoea (rs8048167, rs10654492, rs35087079, rs117686094, rs117260063, rs12231444, rs1491207661, rs11643168, rs12211900 and rs71379941) in predicting **(A)** early onset ≥grade 3 diarrhoea and **(B)** any cycle ≥grade 3 diarrhoea at the respective threshold indicated. ROC analysis results for the top 5 SNPs associated with vomiting (rs193272564, rs140744481, rs200676466, rs4972833 and rs148123087) in predicting **(C)** early onset ≥grade 3 vomiting and **(D)** any cycle ≥grade 3 vomiting at the respective thresholds indicated. The error bars represent the confidence intervals of the curve. Abbreviations: AUC- area under the curve, ROC- receiving operating characteristic curve

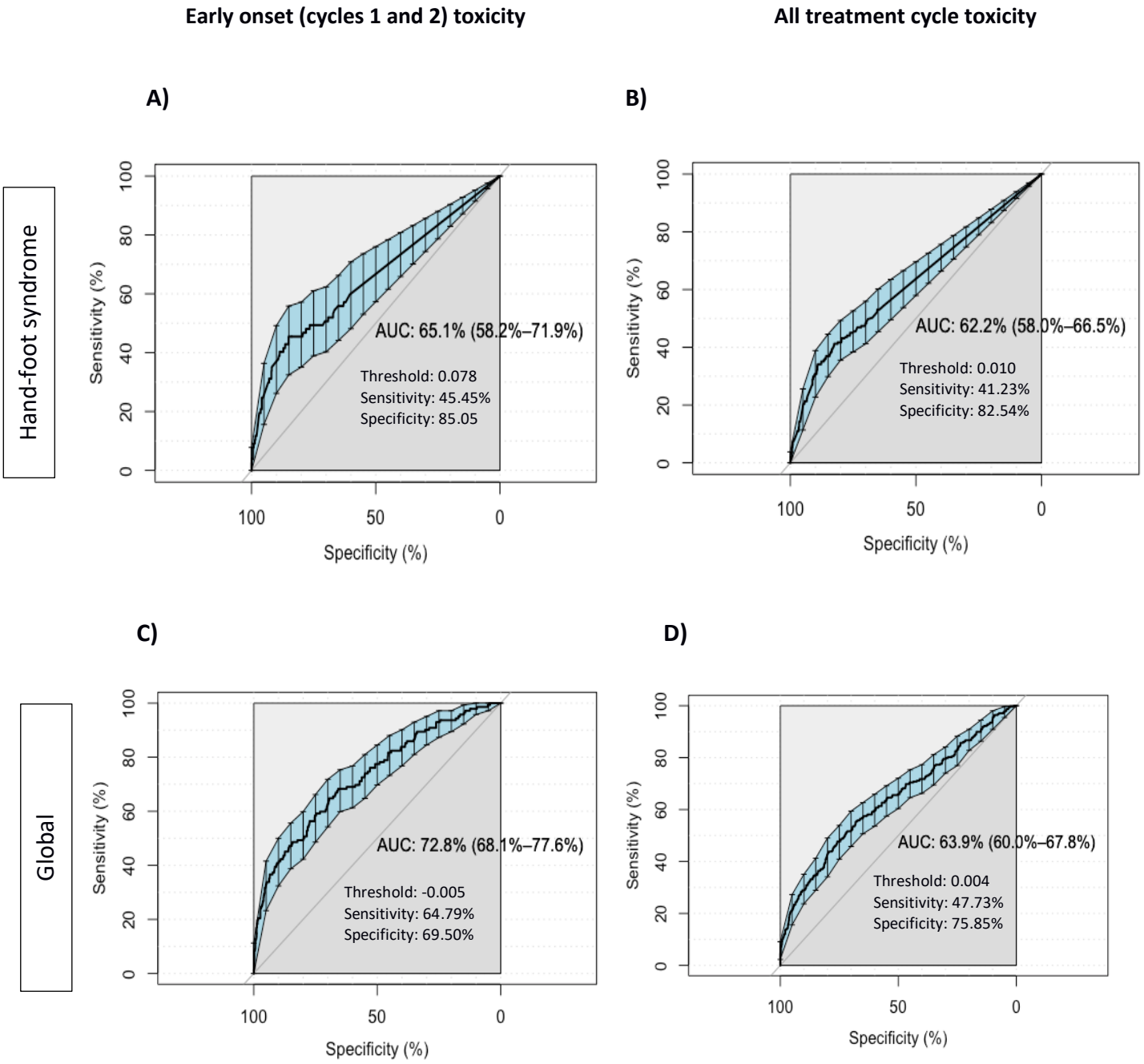


Figure 3.5 AUC-ROC of the top independent SNPS associated with hand-foot-syndrome and global toxicity

ROC analysis results for the top 3 SNPs associated with hand-foot syndrome (HFS) (rs181841887, rs79986886 and rs116448748) in predicting **(A)** early-onset ≥grade 3 HFS and **(B)** any cycle ≥grade 3 HFS at the respective threshold indicated. ROC analysis results for all 26 SNPs associated with early or any cycle ≥grade 3 toxicity of any type (rs8048167, rs10654492, rs35087079, rs117686094, rs117260063, rs12231444, rs1491207661, rs11643168, rs193272564, rs140744481, rs76211390, rs79272399, rs71313929, rs12211900, rs71379941, rs75217739, rs139788123, rs79986886, rs116448748, rs77911328, rs145225093, rs117241924, rs200676466, rs4972833, rs148123087, rs181841887) in predicting **(C)** early onset ≥grade 3 toxicity of any type and **(D)** any cycle ≥grade 3 toxicity of any type at the respective threshold indicated. The error bars represent the confidence intervals of the curve. Abbreviations: AUC- area under the curve, ROC- receiving operating characteristic curve

3.3 Discussion

3.3.1 Previously reported variants of ≥grade 3 toxicity

Genetic variants linked with fluoropyrimidine-based toxicities have been previously identified in candidate gene and pathway analysis. Majority of these toxicity-associated variants have linked with genes encoding for enzymes involved in the fluoropyrimidine metabolism pathway. Examples of these genes include *TYMS/ENOSF1*, *DPYD*, *CDA, CES2* and *MTHFR.* Four *DPYD* risk variants (c.1905+1G>A, c.2846A>T, c.1679T>G and c.1129-5923C>G) are currently being recommended for screening for all patients starting on fluoropyrimidine-based chemotherapy. In addition to identifying novel determinants of capecitabine-induced toxicity, I first assessed the current *DPYD* toxicity risk alleles recommended for testing. I confirmed that these *DPYD* risk variants are associated with fluoropyrimidine-induced toxicity occurring in the first two treatment cycles in QUASAR2. I identified that carriers of any *DPYD* risk alleles had a large increased risk of early-onset ≥grade 3 toxicity (OR=4.05, 95% CI=2.28-7.17, P=5.4 $x10^{-6}$).

3.3.2 Novel variants associated with ≥grade 3 toxicity at genome-wide significance

I have performed the first GWAS of early-onset toxicity following capecitabine treatment. In the GWAS of individual capecitabine toxicities, I compared the GWAS outputs of two methods ("score" and "expected") used to handle genotype uncertainty and found that the "expected" model was more appropriate for detecting risk variants associated with ≥grade 3 toxicities using a binary case-control approach and generated less inflated summary statistics. I performed the GWAS under the genetic additive model which is the model most commonly used GWAS with case-control phenotypes and determines risk by counting the number of genotype risk alleles. Under the "expected" model, only one SNP reached genome-wide

significance in an analysis of the QUASAR2 study (rs181841887, associated with HFS occurring in cycles 1 and 2). A collaborator Katie Watts performed a second GWAS of early-onset toxicity in the COIN trials. She identified two markers reaching GWAS significance, but neither was validated in QUASAR2. The HFS-associated SNP I identified in QUASAR2 could not be validated in the COIN sample due to differences in the genotyping platforms used.

In the meta-analysis of both studies, four novel SNPs were associated with early capecitabine toxicity at GWAS significance (rs117241924, rs140805091, rs146950191, rs2477888). rs140805091, the top SNP associated with early ≥grade 3 haematological toxicity is a significant eQTL of *APP* gene expression in neutrophils and whole blood. *APP* encodes for amyloid beta precursor protein, which is well-known for its pathogenic involvement in the development of Alzheimer's Disease and its effects on neuronal homeostasis and development[212]. However, it is also abundant in non-neuronal cells including neutrophils and is involved in cell proliferation and differentiation[212]. The association between rs140805091 and haematological toxicity has a large effect size (OR= 14.21, 95% CI 5.48-36.86), but further validation is required to assess its translational relevance in clinical practice.

rs2477888 was associated with any ≥grade 3 (global) early toxicity (OR 2.16, 95% CI= 1.64-2.83). This SNP has also been associated with neutrophil counts in a phenome-wide association study (PheWAS) at P= 2.04×10^{-6} [213], which tests associations between a variant and multiple phenotypic traits across large sets of patients as per methods chapter Section 2.3.10[214]. As shown in **Table 3.20**, rs2477888 was associated with early ≥grade 3 HFS, diarrhoea, and haematological toxicities with nominal significance. The other 2 GWAS significant variants have no known associations with treatment-related toxicities warranting further investigation. These four toxicity risk variants identified from the meta-analysis have

not previously been reported in the literature to be associated with capecitabine-induced toxicity.

Table 3.20 Performance of rs2477888 in the meta-analysis of early capecitabine toxicity

Toxicity	OR	Lower CI	Upper CI	P value
Haematological	2.50	1.02	6.10	4.44E-02
Vomiting	2.08	0.90	4.80	8.55E-02
Mucositis	2.02	0.79	5.18	1.42E-01
Diarrhoea	2.02	0.79	5.18	2.18E-03
HFS	2.61	1.72	3.97	6.86E-06

Summary statistics of the SNP in the meta-analysis of COIN and QUASAR2 are shown. Abbreviation: CI- 95% confidence interval, HFS- hand-foot-syndrome, OR- odds ratio

I also identified one marker, rs12211900, associated with any-cycle diarrhoea at genome-wide significance (OR=2.88, 95% CI=1.97-4.19, $P=3.92x10^{-8}$) in the QUASAR2 dataset. This SNP is mapped to an intergenic region with the nearest non-coding RNA genes *LINC02520* and *MIR4462*. Elahimanesh et al. demonstrated that signalling pathways of certain bacterial infections can cause flare-ups of inflammatory bowel disease by interacting with human proteins such as Fibrinogen Gamma Chain and Toll-like Receptor 4 which is affected by microRNAs such as miR-4462[215]. The four novel early toxicity SNPs identified from the meta-analysis and the single any-cycle toxicity SNP detected from QUASAR2 that reached GWAS significance will need to be validated in an external cohort. Upon validation, further pathway and functional analysis will be required to establish the biological impact and clinical utility for predicting toxicity of these variants.

MIR27A variants linked with fluoropyrimidine toxicity

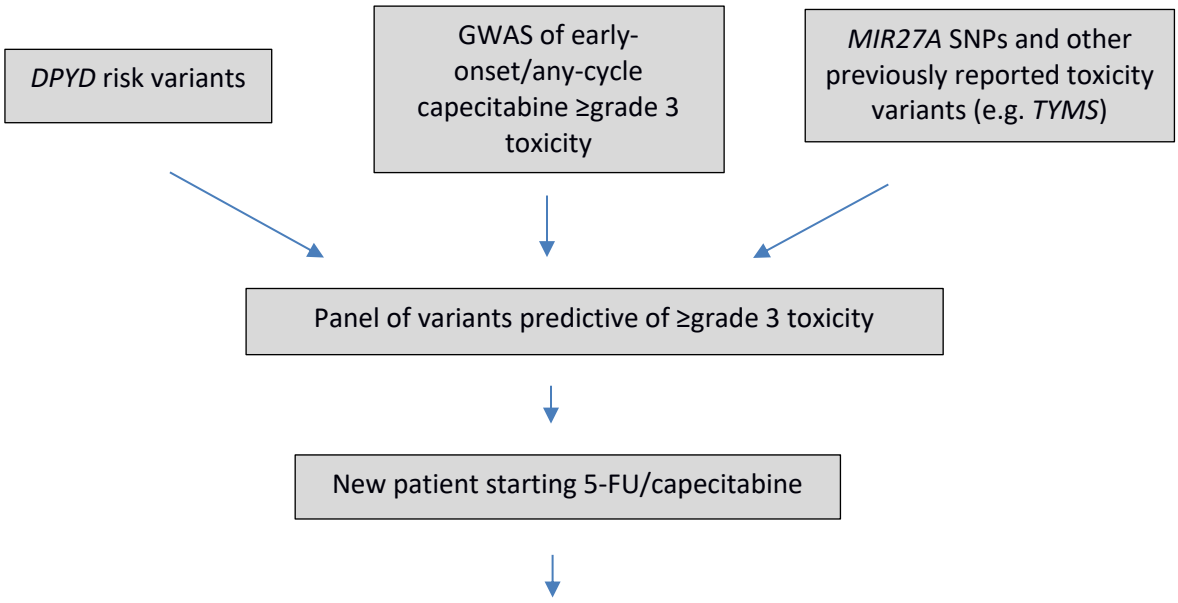
MiR-27A are microRNAs involved in the post-transcriptional regulation of *DPYD* and have been associated with decreased DPD activity. Two *MIR27A* variants, rs895819 and rs11671784 have been associated with risk of fluoropyrimidine toxicity. I investigated the association of these

two *MIR27A* variants in datasets from the QUASAR2 trial and a Dutch cohort. In a meta-analysis of QUASAR2 and the Dutch cohort, I showed that rs895819 was significantly associated with early-onset ≥grade 3 diarrhoea, GI toxicities and global toxicity. This is the first time a significant association has been reported for this SNP in patients not stratified by *DPYD* low or no function carrier status. The meta-analysis did not identify an association between rs11671784 and ≥grade 3 early toxicity. However, larger sample sizes will be required to completely rule out an effect given the lower MAF of this SNP.

The effect of rs895819 was strongest under the recessive model, suggesting that homozygous carriers of the toxicity-associated 'C' allele are at the highest risk of toxicity. However, even under a recessive model, the effect of carrying two risk alleles is modest (OR=1.43; observed in patients treated with capecitabine alone) compared to the risk associated with carrying a single *DPYD* low or no function allele (risk of early ≥grade 3 toxicity in carriers is ~4). It is difficult to know whether predictive screening should include this SNP, particularly given the conflicting results for *DPYD* low or no-function allele carriers. Previous reports found a higher risk of toxicity in carriers of *DPYD* low or no function allele who also carried one or more rs895819 'C' allele[115]. I did not detect an increased risk of ≥grade 3 toxicity in carriers of rs895819 'C' allele and a *DPYD* low or no- function allele compared to those carrying a *DPYD* low or no- function allele and no rs895819 'C' alleles. Indeed, adding rs895819 into a test with the CPIC *DPYD* risk alleles does not improve the sensitivity of the test (both 10%) with only a slight improvement in specificity (95% vs 96% when considering rs895819 risk allele carriers). Prediction of individual toxicities based on the AUC was also similar with or without adding rs895819. It may be that rs895819 has utility in a stepwise testing process where carriers who are not *DPYD* low or no risk carriers can then be tested for rs895819 to predict the risk of GI toxicity. Given the modest effect size on the risk of GI toxicity (OR<2) and the low specificity (30%) of the test, including rs895819 testing for this toxicity is unlikely to be clinically justified.

3.3.3 Determining toxicity risk using polygenic risk scores and toxicity prediction models

I generated per patient polygenic risk scores for QUASAR2 participants using the effect size estimated multiplied by the number of risk alleles the patient carried. As QUASAR2 contributed to the meta-analysis this approach is biased. Still, in the absence of an independent dataset it demonstrates the performance of the identified sets when used in combination, with the caveat that the true sensitivity and specificity of tests including these SNPs need to be assessed in independent sample sets. The AUCs of models including SNPs associated with individual toxicities (e.g. mucositis) were high (>0.8) for the prediction of the corresponding toxicity, but the model including all top 26 SNPs had a suboptimal AUC of 0.6 for predicting any ≥grade 3 toxicity across all treatment cycles. This difference in AUC could be because SNPs that are risk alleles for one toxicity are not risk alleles for another, so when combined they cannot correctly identify those at risk of any ≥grade 3 toxicity. It may be that stepwise testing estimating each toxicity variable and then a summary measure of toxicity risk based on the results at each step provide better risk prediction. An alternative approach would be to perform upfront testing of all known toxicity-associated risk variants and a personalised risk profile is generated for each patient depending on which toxicity risk variants they are carriers of. Recommendations on prevention strategies are then tailored to each patient depending on their risk. An example of this approach is demonstrated in **Figure 3.6**.



Patient	Haem	Diarrhoea	Mucositis	HFS	Vomiting	Recommendations
A	High	High	Low	Low	Low	Dose reduction
B	Low	Low	Low	High	High	Topical emollients, prophylactic antiemetics
C	Low	High	High	Low	Low	Dose reduction, anti-inflammatory mouthwashes

Figure 3.6 Possible predictive model of ≥grade 3 toxicity risk in patients starting fluoropyrimidine-based therapies

Patients receive upfront testing of previously reported and novel toxicity risk variants before commencing 5-FU/capecitabine chemotherapy. Novel toxicity risk variants detected from the GWAS will require further validation. Eg Patient A is found to be a carrier of variants predictive of ≥grade 3 haematological and diarrhoea toxicity but does not carry risk alleles of variants predictive of other toxicities. A dose reduction for his chemotherapy is recommended to prevent the occurrence of ≥grade 3 neutropenic and colitis events.

3.3.4 Limitations of analyses

~15% of patients in QUASAR2 experienced a ≥grade 3 capecitabine-induced toxicity in the first two treatment cycles. There were only 141 cases in the analysis, which is very modest by GWAS standards. This modest number of cases limits the power of the QUASAR2 study to detect the full range of genetic variants associated with risk of developing toxicity. QUASAR2 was only powered to detect variants with large effect sizes. The meta-analysis combining the COIN and QUASAR2 datasets was conducted to help increase the sample size and achieve a more powered analysis. However, some differences exist between the two trials, including the

type of genotyping arrays, association analysis tools, cancer staging of patients and treatment regimes received may have impacted the results and limited the number of variants that reached statistical significance. For example, patients in arm B of QUASAR2 also received bevacizumab, and COIN included patients on oxaliplatin. The dose of capecitabine received by patients in QUASAR2 in the adjuvant setting is higher compared to that received by patients in COIN (1250mg/m^2 versus 1000mg/m^2 when used in combination with oxaliplatin in COIN). Some SNPs present in either study alone look promising, e.g. rs179893 (7:106030816_C_CA) and rs200507264 (7:106030823_C_CA), which are both associated with ≥grade 3 haematological toxicity in COIN. However, I could not perform further genotyping to confirm their predictability. It was not possible to confirm whether these SNPs are in linkage disequilibrium using 1000G and UK Biobank data as these markers are rare (MAF ~ 2%). They also map to a unique region of short interspersed nuclear elements (SINE) with repeat runs of 'CA' nucleotide bases, which may pose a challenge for designing genotyping assays for these two SNPs. The top variants identified from the GWAS would require further testing in an independent cohort to validate their association with ≥grade 3 capecitabine toxicity.

3.3.5 Implication of findings

The ability to detect early-onset toxicity has important translational relevance in clinical practice. Identifying patients at high risk of experiencing ≥grade 3 toxicities early in their treatment pathway can help guide personalised treatment modifications or prevention measures to minimise the toxicity burden. Together with the currently tested four *DPYD* risk variants, the top SNPs identified from my GWAS of early capecitabine toxicity and the *MIR27A* rs895819 SNP if validated, could potentially contribute to a panel of candidate genetic polymorphisms to identify patients at risk of ≥grade 3 toxicity. As discussed above, there may

be a need to generate risk predictions for specific toxicities and an overall summary. Alongside the appropriate dose reductions, simple intervention measures could be used to prevent and monitor these toxicities.

CHAPTER 4: FUNCTIONAL ANALYSIS OF A GENE ASSOCIATED WITH 5-FU INDUCED CARDIOVASCULAR TOXICITY AND IDENTIFICATION OF GENETIC DETERMINANTS OF CARDIAC TOXICITY

4.1 Introduction

Fluoropyrimidine-induced cardiotoxicity can occur in up to 35-68% of patients treated with capecitabine and intravenous 5-FU respectively[21]. Cardiac toxicity can be severe and potentially life-threatening. According to a systematic review of 30 studies, serious cardiac events occurred in up to 2% of patients with mortality rates ranging from 0-8%[216]. No clear associations have been found between pre-existing cardiac disease or known risk factors for cardiac disease such as hypertension, hyperlipidaemia, obesity and smoking and the development of 5-FU associated cardiotoxicity[216]. The pathophysiology behind the development of cardiac toxicity in patients receiving 5-FU chemotherapy is also not established.

4.1.1 Genetic markers of fluoropyrimidine-induced cardiotoxicity

Genetic variation is a potential determinant of cardiotoxicity, given the variation in inter-patient susceptibility to developing cardiotoxicity. This discrepancy has been observed in patients receiving anthracyclines where the risk of cardiotoxicity is not strictly dose-dependent and cardiomyopathy has been reported in patients receiving low doses of anthracycline, below the recommended dose threshold[217]. There are currently no known genetic factors that predispose towards fluoropyrimidine-induced cardiotoxicity. Prior association analyses conducted by our lab group identified a cluster of variants in strong linkage disequilibrium (LD) with one another on chromosome 14 that are strongly associated with the presence of cardiac toxicity in the QUASAR2 clinical trial[159](**Figure 4.1**)(manuscript in preparation).

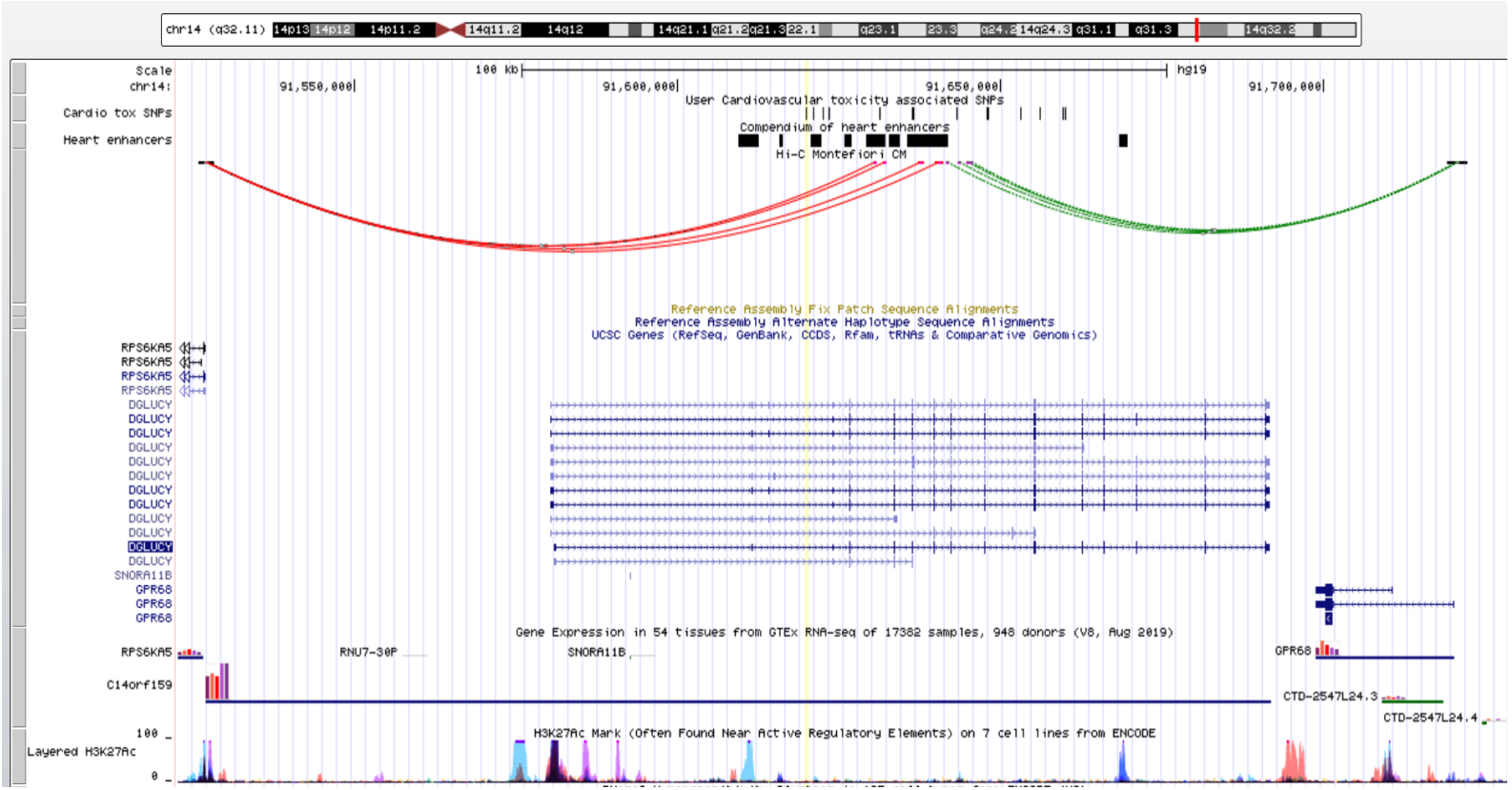


Figure 4.1 Genetic variants associated with cardiovascular toxicity

Chromosome positions of the cardiovascular toxicity associated SNPs identified in the QUASAR2 analysis. Chromosome positions of enhancers identified by Dickel et al. and Hi-C promoter contacts reported by Montefiori et al. were uploaded to the UCSC genome browser as a custom track[184][218][219]. Whilst the cardiovascular toxicity SNP maps to *DGLUCY,* there is some evidence of interaction between the region containing the cardiovascular toxicity associated SNPs and the promoters of nearby genes in cardiomyocytes.

Most of the cluster of variants map to an intron of the *DGLUCY* gene, which until recently was annotated as C14orf159. Other nearby genes are *RPS6KA5* and *GPR68*. Two of the SNPs associated with cardiac toxicity mapped to the coding regions of *DGLUCY.* rs4900071 is a synonymous variant, whereas rs4900072 is identified as a missense variant in some transcripts of *DGLUCY* and a splice donor variant in other transcripts. rs4900071 and rs4900072 overlap with the promoter histone mark H3K4me3 in skeletal muscle, suggesting that these SNPs may affect transcription factor binding and gene expression. Most of the cardiac toxicity SNPs were found in regions that overlap with enhancer histone marks and three SNPs rs4904753, rs4904755, and rs4900071, are found in open chromatin regions. rs4904755 is also found in a region bound by the protein oestrogen receptor alpha (Erα) and has a RegulomeDB score of 2b. rs4900072 in linkage disequilibrium (LD) with rs10148571 has also been shown to be associated with levels of glutamate in the urine of patients with reduced kidney function [220].

The cardiotoxicity-associated alleles of the lead variant (rs4904753) are associated with increased levels of *DGLUCY* gene expression in skeletal muscle ($P=2 \times 10^{-8}$) and decreased expression of *DGLUCY* in heart atrial appendage ($P=1.3 \times 10^{-6}$). Cancer MRes student Cameron Campbell (thesis title: Capecitabine-induced cardiovascular toxicity) performed a transcript level analysis of the 35 transcripts of *DGLUCY* and identified that the risk alleles of the cardiac toxicity associated SNPs are associated with decreased expression of ENST00000256324.10 in all tissue types examined but increased expression of ENST00000428926.2 in multiple tissues. ENST00000256324.10 and ENST00000428926.2 both contain 14 exons and encode two different protein isoforms of *DGLUCY*: isoform 2 (Ensembl canonical isoform) and isoform 1, respectively. These two isoforms have the same domain structure, including two domains of unknown function (DUF), DUF4392 and DUF1445.

Given the promising outputs based on the QUASAR2 dataset alone, we were motivated to expand our GWAS cohort and, hence in collaboration with groups from the COIN and SCOT

study, I was able to perform an agnostic meta-analysis GWAS combining datasets for all three trials to identify novel genetic markers predisposing towards cardiotoxicity.

4.1.2 Pre-clinical studies determining the role of DGLUCY

The *DGLUCY* gene encodes for the D-glutamate cyclase protein which is predicted to be involved in glutamate metabolism. The impact of DGLUCY on the heart was analysed in a preclinical study by Ariyoshi et al. who induced heart failure in transverse aortic constriction-operated mouse hearts and examined changes in protein expression[187]. The protein with the largest reduction in expression was the mouse homologue of C14orf159, also known as 9030617O03Rik. This paper showed sequence similarity between the mammalian 9030617O03Rik and bacterial aspartate/glutamate/hydantoin racemase and demonstrated that 9030617O03Rik has D-glutamate cyclase (DGLUCY) activity and converts D-glutamate to 5-oxo-D-proline[187]. Downregulation of DGLUCY in DGLUCY-deficient mice lead to an increase in D-glutamate levels resulting in onset of heart failure. The physiological role of D-glutamate is not yet well defined, however, findings from this study indicate that D-glutamate could play an important role in cardiac function via metabolism of DGLUCY[187].

Expression of DGLUCY is highest in the heart, followed by the kidney and liver but has low expression in other tissues. Ariyoshi et al also examined the cardiac localization of 9030617O03Rik by separating myocytes and fibroblasts from rat hearts and found that the protein is predominantly present in cardiomyocytes. Further cell fractionation analysis showed that the protein is located in the mitochondria matrix[187]. Katane et al. further examined the structural and enzymatic properties of DGLUCY. They worked with purified recombinant mouse DGLUCY protein and confirmed that the protein was stereospecific and

only catalyses the conversion of D-glutamate to 5-oxo-D proline with no effect on the conversion of L-glutamate to 5-oxo-L proline[221]. The authors also identified a preference for the enzyme to convert D-glutamate to 5-oxo D proline over the conversion of 5 oxo-D-proline to D-glutamate. They also suggest that DGLUCY is a metalloenzyme with its activity impacted by several cations- $MnCl_2$, ATP, ADP and AMP.

4.1.3 Effects of 5-FU on amino acid metabolism

Given the potential adverse effects of D-glutamate on cardiac function demonstrated by Ariyoshi et al, it would be interesting to assess if other amino acids could also contribute to cardiac toxicity induced by 5-FU.

Some studies have also assessed the impact of 5-FU on amino acid metabolism and the tricarboxylic acid (TCA) cycle. To evaluate the effects of chemotherapy by metabolomic studies, Tomita et al. quantified the concentration of 21 amino acids in human colorectal adenocarcinoma cell culture treated with chemotherapy using fluorescence derivatization-HPLC[222]. The authors found significant changes in the measured concentrations of all 21 amino acids, especially serine and aspartic acid in cells treated with 5-FU compared to untreated cells[222]. 5-FU catabolism generates metabolites such as fluoroacetate, which causes inhibition of the TCA cycle enzyme aconitase[223]. Using the human mitochondrion model of cardiomyocytes, Lischke et al. found effects of ATP depletion and citrate accumulation as the result of aconitase inhibition which can contribute to 5-FU cardiotoxicity[224]. Increased uptakes of valine, arginine, proline and glutamate effectively compensate for the impaired energy metabolism[224].

4.1.4 Investigating the role of *DGLUCY* on cardiac toxicity in the presence of 5-FU

All the experiments done so far to further understand the role of 9030617O03Rik/*DGLUCY* have been done without 5-FU, where downregulation of DGLUCY was associated with increased D-glutamate levels. We wanted to investigate the impact of DLGUCY and changes in D-glutamate levels in the presence of capecitabine or 5-FU. This will help us further understand the role of D-glutamate on cardiac function and fluoropyrimidine-related cardiotoxicity. We worked with mouse cardiomyocyte cell line HL-1 given that all the work that had previously been done on *DGLUCY* had been done in mice, and we expected to need large numbers of cells for the detection of D-glutamate, previously shown to be considerably less abundant than its enantiomer L-glutamate. HL-1 cells can also be passaged repeatedly and retain the ability to contract and resemble the phenotypic characteristics of adult cardiomyocytes[225]. Reverse transcriptase-PCR showed that their gene expression pattern is similar to adult atria myocytes[225]. The cells were a kind gift from Katja Gremlich (University of Birmingham). A MRes student in our group investigated the effect of treatment with increasing concentration of capecitabine on the expression of *DGLUCY* and nearby gene *RPS6KA5* in HL-1 cells. Results from qPCR measurements showed no significant impact after 24 hours of treatment with Capecitabine on *DGLUCY* and *RPS6KA5* gene expression.

Ariyoshi et al. used two-dimensional high-performance liquid chromatography (2D-HPLC) to analyse the amino acid profiles including L and D-enantiomers, in 9030617O03Rik-deficient mouse hearts[187]. The L-glutamate: D-glutamate ratio was approximately 850:1 in wild-type mouse hearts and 140:1 in 9030617O03Rik deficient mouse[187]. In this analysis, in collaboration with the Institute of Metabolism and Systems Research, we used a gas chromatography-mass spectrometry (GC-MS) column specifically designed to detect D- and L-enantiomers to detect levels of both D-glutamate and L-glutamate in mice cardiomyocytes.

4.1.5 Chapter aims

This chapter had the following aims:

1. Use functional analyses to define the role of *DGLUCY* in cardiomyocyte viability and metabolism under 5-FU treatment
2. Determine whether *DPYD* variants currently screened for clinically to predict those at risk of fluoropyrimidine induced toxicity are associated with cardiovascular toxicity
3. Identify novel genetic determinants of fluoropyrimidine induced cardiotoxicity by combining datasets from QUASAR2, SCOT AND COIN trials

4.2 Results

4.2.1 Functional analysis of *DGLUCY* gene

4.2.1.1 Effects of 5-FU chemotherapy on viability of HL-1 cells

To determine the correct concentration of 5-FU to treat HL-1 cells, a 96-well plate was seeded with 5000 cells per well and treated with different concentrations of 5-FU chemotherapy (1µM, 5µM, 10µM, 25µM, 100µM and 200µM). This range of concentrations was selected as similar 5-FU concentrations to treat colorectal cancer cell lines were attempted by the Beggs group in our institute. Tawfik et al. also constructed 5-FU dose-response curves using MTT assays in their experiments on colon cancer cells and found the calculated IC_{50} to be 185µM after 1 day and 1.48µM after 5 days[186]. Five technical replicates of the different 5-FU drug concentrations were tested. ~24 hours after the first treatment dose, MTS assay reagent was added to each well to measure cell viability. The viability of the cells treated with different drug concentrations with an incubation time of 15 mins, one hour and four hours are shown in **Figure 4.2**. Absorbance from the cells was highest after four hours of incubation with MTS.

Untreated cells as controls (0uM) have the highest absorbance values indicating higher cell viability than those treated with 5-FU chemotherapy. MTS incubation time and 5-FU dose concentration were significantly associated with different absorbance values ($P=2x10^{-16}$ and P=0.009 respectively). However, following multiple pairwise comparisons of the different drug doses, significant mean differences were only detected between the untreated group and treatment with drug doses of 10µM (P=0.007) and 100µM (P=0.02).

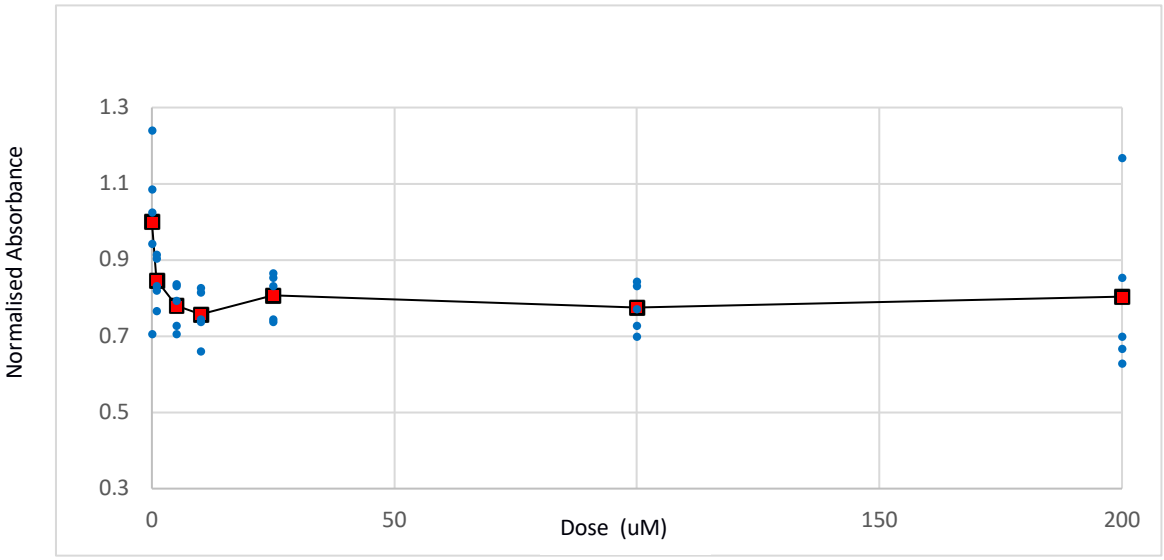


Figure 4.2 Line plots of HL-1 cells treated with a single dose of 5-FU chemotherapy following MTS assay administration.

Six different 5-FU drug concentrations (1µM, 5µM, 10µM, 25µM, 100µM and 200µM) were tested with untreated cells as negative controls (0µM). Cells were treated for 24 hours before MTS assays were performed. The x-axis represents the different 5-FU drug doses used to treat the cells. The y-axis starts from 0.3 and represents the absorbance levels of the MTS assay detected, normalised to untreated cells at baseline (mean value normalised to 1). The time taken from initial MTS assay is 4 hours. The blue dots represent each absorbance value and the red squares represent the mean absorbance value for the individual doses. Statistical analysis was performed using two-way ANOVA and Tukey HSD test for multiple testing.

I also tried a different dosing schedule whereby cells were treated twice three days apart with the same concentration of 5-FU both times. ~24 hours after the second dose was given, MTS assay absorbance for each well treated with the different drug concentrations was measured and the results are presented in **Figure 4.3**. The maximum concentration tested this time was increased to 400µM. Similar to the first experiment **(Figure 4.2)**, the untreated cells (0uM)

showed the highest absorbance at four hours post-incubation with MTS reagent, which was significantly different to the treated cells following multiple pairwise comparisons testing. Even more so than in the first experiment, there was no difference in absorbance values when comparing different concentrations of 5-FU (P>0.99). This observation suggests that 5µM of 5-FU decreases the viability of the cells, but higher concentrations do not have an additional impact on cell viability.

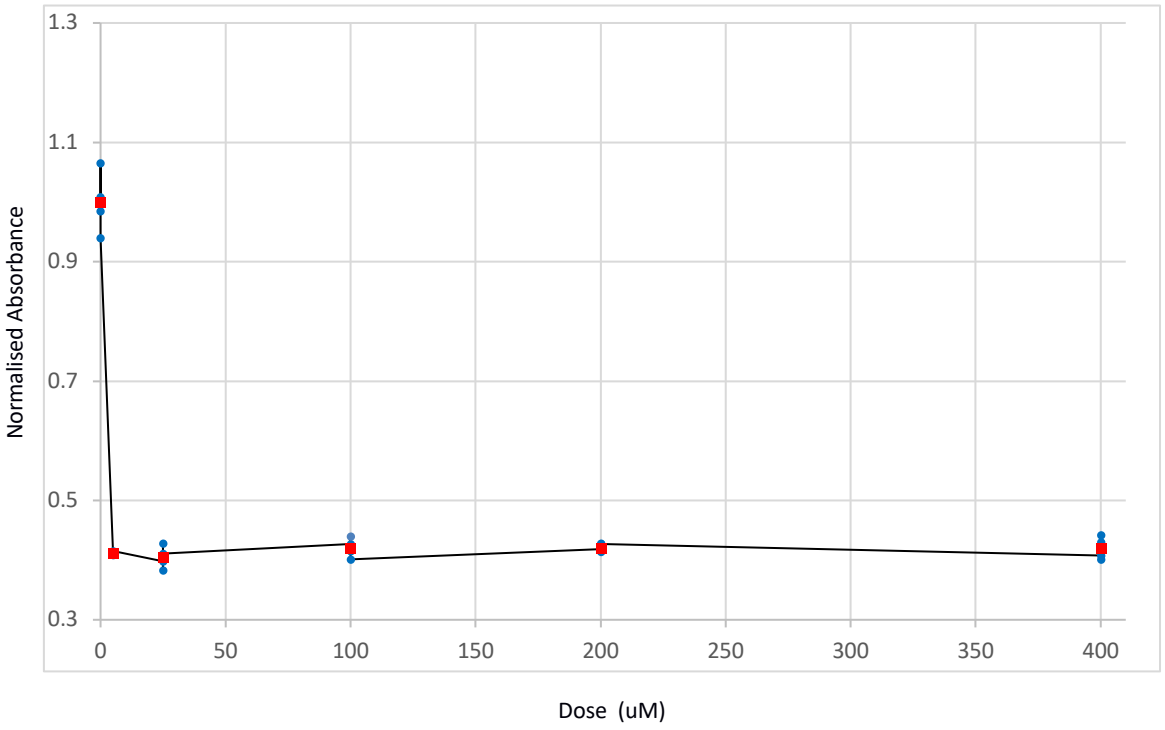


Figure 4.3 Line plots of HL-1 cells treated with two doses of 5-FU chemotherapy following MTS assay administration.

The cells were treated twice for each drug concentration (5µM, 25µM, 100µM, 200µM and 400µM) on day 1 and day 4 from when the cells were first plated. Untreated cells served as the negative control (0µM). The x-axis represents the different 5-FU drug doses used to treat the cells. The y-axis starts from 0.3 and represents the absorbance levels of the MTS assay detected, normalised to untreated cells at baseline (mean value normalised to 1). The time taken from initial MTS assay administration to measurement of the absorbance results is 4 hours. The blue dots represent each absorbance value and the red squares represent the mean absorbance value for the individual doses.

4.2.1.2 Gene expression of DGLUCY following treatment with 5-FU

Having established that higher concentrations of 5-FU did not decrease cell viability further, I treated HL-1 cells with a range of lower 5-FU drug concentrations (0.1µM, 0.25µM, 1µM, 5µM and 25µM) with untreated cells acting as negative controls. The aim was to determine what impact exposure to the drug had on expression of genes in the locus associated with 5-FU induced cardiovascular toxicity. Following 24 hours of treatment, RNA extraction and conversion to cDNA were performed, and the gene expression of *DGLUCY*, *RPS6KA5* and two housekeeping genes *RPL32* and *PGK1* were measured via qPCR. Two technical replicates were analysed for each 5-FU concentration tested. There was minimal differences between the expression of DGLUCY in untreated and treated cells (Figure 4.4), represented by the delta-delta CT value of DGLUCY of -0.10 in cells treated with 5-FU. The mean Ct value of target gene *DGLUCY* in untreated cells was 24.08 and in cells treated with 0.1µM of 5-FU, the mean Ct value was similar (23.99).

Similarly, there was no change in expression of the second target gene *RSP6KA5* when untreated and treated cells were compared. Ct values for both genes were undetermined when cells were treated with higher concentrations of 5-FU. I, therefore, do not provide any evidence of expression levels of *DGLUCY* or *RPS6KA5* changing upon treatment with 5-FU.

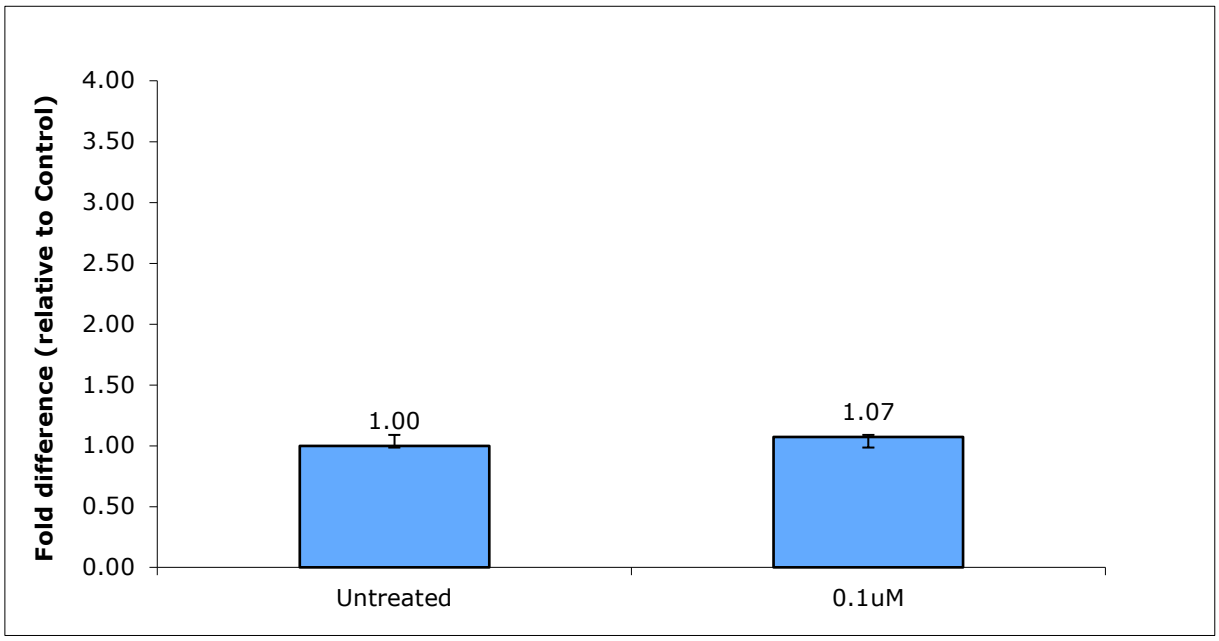


Figure 4.4 Fold difference in gene expression between untreated and treated HL-1 cells.
The x-axis represents the different cell conditions (untreated or treated with 0.1µM of 5-FU). The y-axis represents the fold difference in gene expression of both cell conditions after considering the Ct value comparisons for the housekeeping gene *PGK1* and target gene *DGLUCY* with untreated cells as the control group (set to 1.00 as a constant). Error bars indicate the standard deviation.

4.2.1.3 Detection of L-glutamate and D-glutamate by gas chromatography-mass spectrometry (GC-MS)

Accepting that there is still some uncertainty regarding the target gene in the locus associated with 5-FU induced cardiovascular toxicity, I began a collaboration with Jennie Roberts and Dan Tennant in the Institute of Metabolism and Systems Research (UoB) to set up gas chromatography-mass spectrometry to measure levels of D-glutamate. Chiral GC columns were purchased from Agilent and were tested for their ability to detect synthetic L-glutamate and D-glutamate. As can be seen from **Figure 4.5** below, the two enantiomers can be resolved. The peaks at 25 minutes are L and D glutamate, and the much smaller peak at 24 minutes is pyroglutamate. Glutamate can spontaneously cyclise to pyroglutamate in vitro.

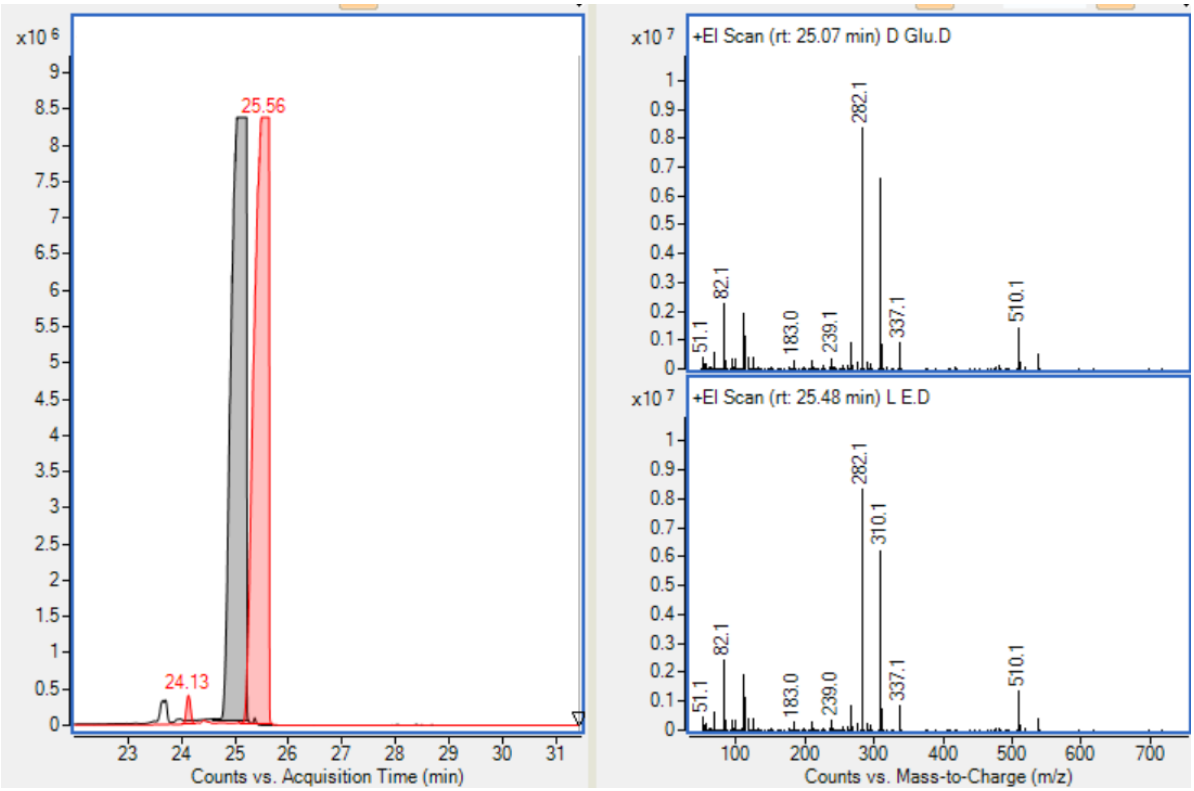


Figure 4.5 Chiral glutamate on Agilent Chirasil-Val

D/L glutamate were fully resolved with main chromatographic peak at RT 25 minutes and characteristic mass-to-charge (m/z 282 and 310). Note small peaks at room temperature at 24 minutes are pyroglutamate. The column was also shown to resolve L and D pyroglutamate.

We then determined whether D-glutamate was detectable in HL-1 cells and mouse heart tissue. We only had access to wild-type mice but used the siRNA described above to knock down *DGLUCY* in HL-1 cells. Using the limited number of heart samples we could process, we could not optimise the tissue extraction sufficiently to prevent glutamate from degrading to pyroglutamate during sample processing. Homogenising the tissue sample with the Precellys tissue homogeniser requires friction which produces heat. We could only detect L-pyroglutamate in our tested samples **(Figure 4.6).**

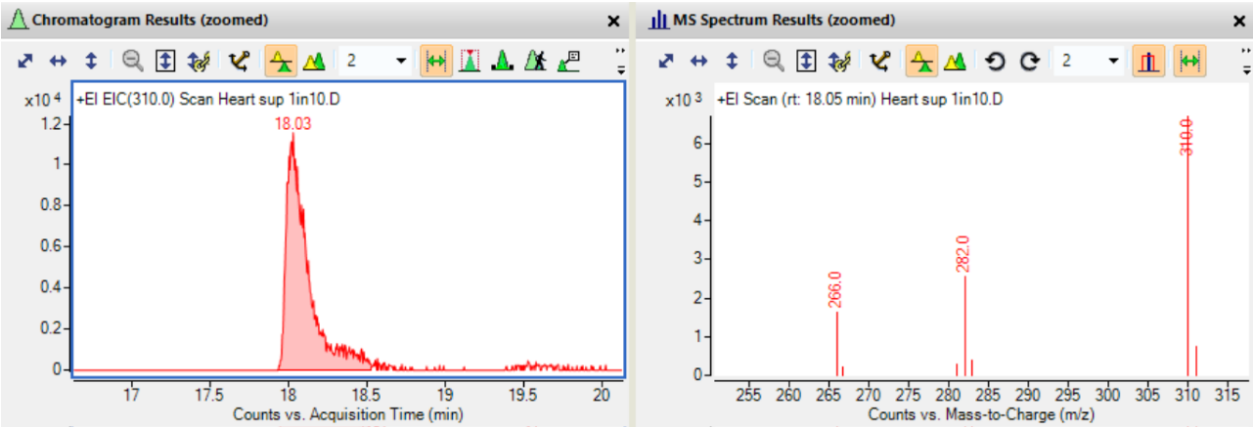


Figure 4.6 GC-MS chromatogram showing the detection of L-pyroglutamate from mouse heart tissue
L-pyroglutamate was detected at m/z 310 and m/z 282 with the main chromatographic peak at RT 18 minutes from tissue samples. Image obtained courtesy of Jennie Roberts.

HL-1 mouse cardiomyocytes for analysis by mass spectrometry were plated in triplicates of 50,000, 100,000 and 200,000 cells. The cells were harvested as per methods chapter Section 2.6.4. A large amount of L-glutamate was detected in the triplicate of 200,000 HL-1 cells and trying to extrapolate from the information provided by Ariyoshi et al., we think we should have just been able to detect D-glutamate if it had been present, however, only L-glutamate was detectable **(Figure 4.7)**.

L-glutamate RT 25.12min

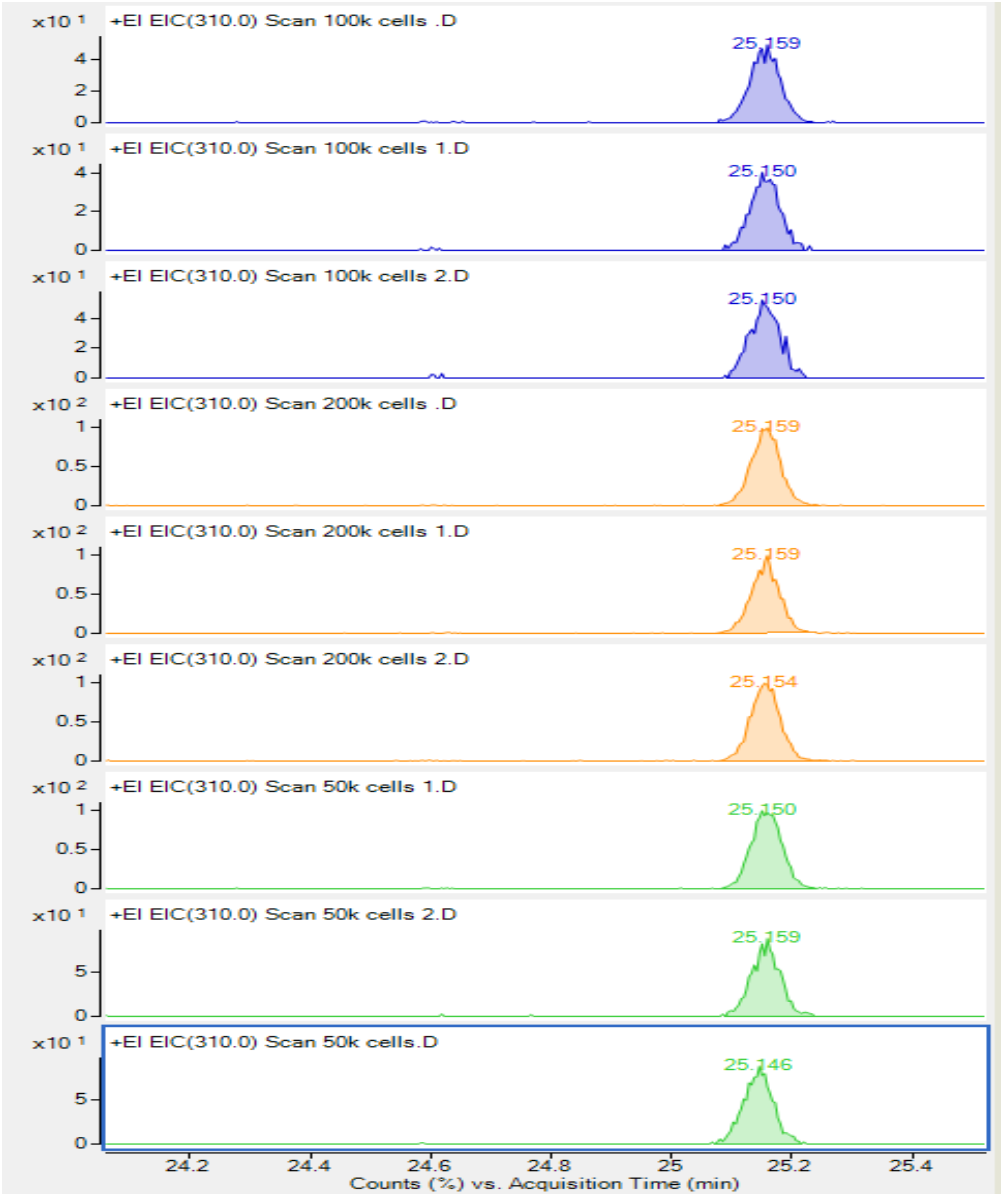


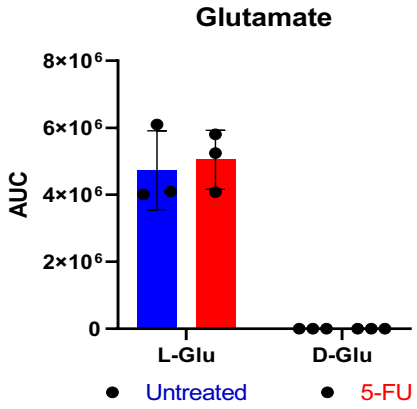
Figure 4.7 GC-MS chromatogram showing the detection of the L-glutamate extracted from HL-1 mouse cardiomyocytes.

Peak areas of L-glutamate extracted from triplicates of 50,000 (green), 100,000 (blue) and 200,000 (orange) HL-1 cells. The x-axis shows the RT for L-glutamate (~25.12 minutes), and the y-axis corresponds to the L-glutamate ion counts (%). No peak is visible at RT 24 mins indicating the absence of D-glutamate. Image obtained courtesy of Jennie Roberts.

4.2.1.4 D-glutamate is undetectable after treatment of HL-1 cells with 5-FU

As D-glutamate was not detectable from cell counts of 200,000 cells, higher triplicates of 500,000 cells, both untreated and treated with 0.1µM of 5-FU were harvested for metabolite extraction of glutamate. Based on the gene expression analysis results in Section 4.2.2, 0.1µM was the selected drug concentration. Similar to the previous GC-MS results, only the L-glutamate enantiomer was detectable in untreated cells and cells treated with 5-FU. D-glutamate was not detectable as shown in **Figure 4.8.**

a)



b)

Whole Cell Extracts	L-isomer		D-isomer	
AUC	**Untreated (Control)**	**5-FU treated**	**Untreated (Control)**	**5-FU treated**
Glutamate	4732685 ± 1180219	5040907 ± 879793	nd	nd

Figure 4.8 Comparison between L-glutamate and D-glutamate levels between untreated and 5-FU treated HL-1 cells

a) Area under the curve (AUC) levels for L- and D-glutamate in both untreated (blue bars) and 5-FU treated cells (red bars). Experiments were performed in triplicate with error bars shown. b) AUC quantities with standard deviations of the L and D isomers are shown. Abbreviations: nd- not detectable. Image obtained courtesy of Jennie Roberts.

4.2.1.5 Knockdown of DGLUCY by siRNA transfection in HL-1 cells

It was reported by Ariyoshi et al. that the D-glutamate: L-glutamate ratio is approximately 1:850 in wild-type mice and 1:140 in *DGLUCY* knock-out mice. With the hypothesis that

knockdown of *DGLUCY* might be required to detect D-glutamate using mass spectroscopy, I obtained siRNA targeting *DGLUCY* to mimic *DGLUCY* gene knockdown. 'Scramble' (SCR) siRNA was used as a transfection control for comparison. Two different concentrations of siRNA (30pmol and 60pmol) were tested. As shown in **Figure 4.9**, efficient knockdown of DGLUCY was achieved using the 30pmol siRNA pool, and the knockdown effect was largest 48 hours post-transfection. Knockdown efficiency was assessed using RT-PCR.

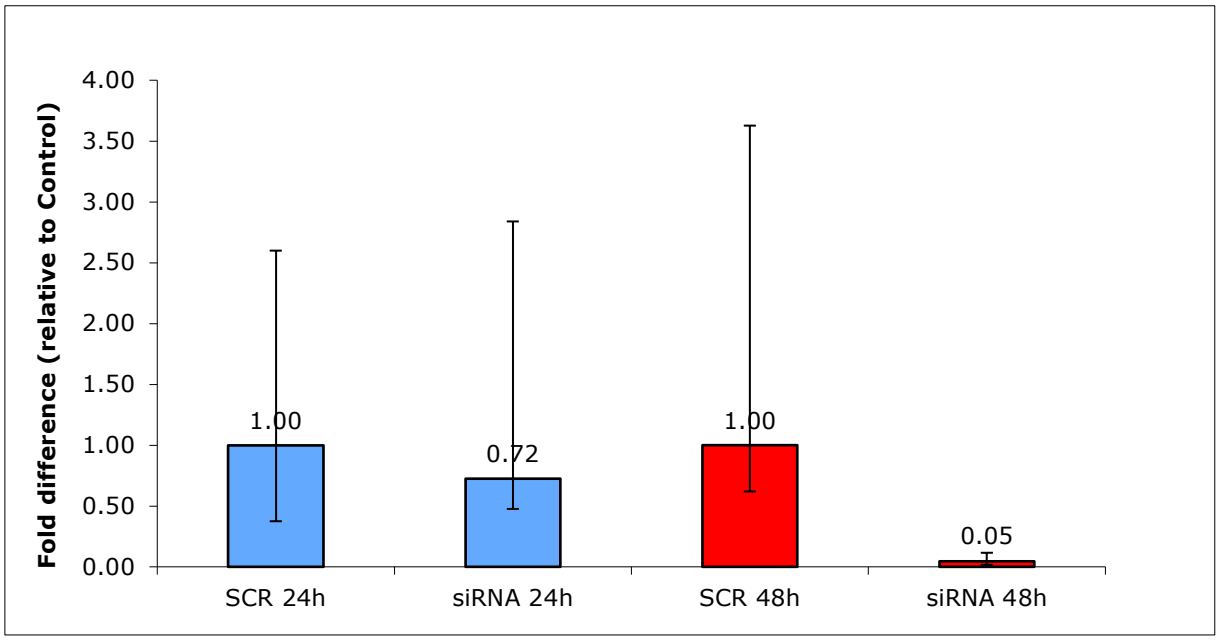


Figure 4.9 Fold difference in gene expression between the SCR cells and cells transfected with 30pmol siRNA targeting *DGLUCY* at 24 and 48 hours post-transfection.
The fold difference in *DGLUCY* gene expression in the siRNA knockdown cells relative to the scramble (SCR) treated cells is shown. Ct values were first normalised using the Ct values of housekeeping gene *PGK1.* Two technical replicates were included for 24 and 48-hour periods of transfection testing. Error bars indicate the 95% confidence intervals.

4.2.1.6 D-glutamate was still undetected after siRNA knockdown of DGLUCY in either mitochondrial or cytoplasmic cell fractions

Given the difficulties experienced detecting D-glutamate using wild-type cells, I decided to knock down *DGLUCY* before preparing samples for GC-MS. As detailed in Section 4.2.5, I treated HL-1 cells with *DGLUCY* or Scramble siRNAs. Untreated cells were also prepared alongside these. I then prepared whole cell extracts for GC-MS or isolated mitochondria using

the methods described in Chapter 2: Section 2.7**.** 9030617O03Rik/DGLUCY is reported to be a mitochondrial protein, so we wanted to test whether high levels of L and D glutamate could be detected in mitochondria. Concentrations of digitonin ranging from 25ugml – 300ug/ml were tested to assess the lowest concentration required to permeabilise the cells while leaving the mitochondria intact. As shown in **Figure 4.10**, HL-1 cells were already stained with trypan blue after adding 25ug/ml of digitonin, indicating that this concentration is sufficient to enable cell permeabilisation.

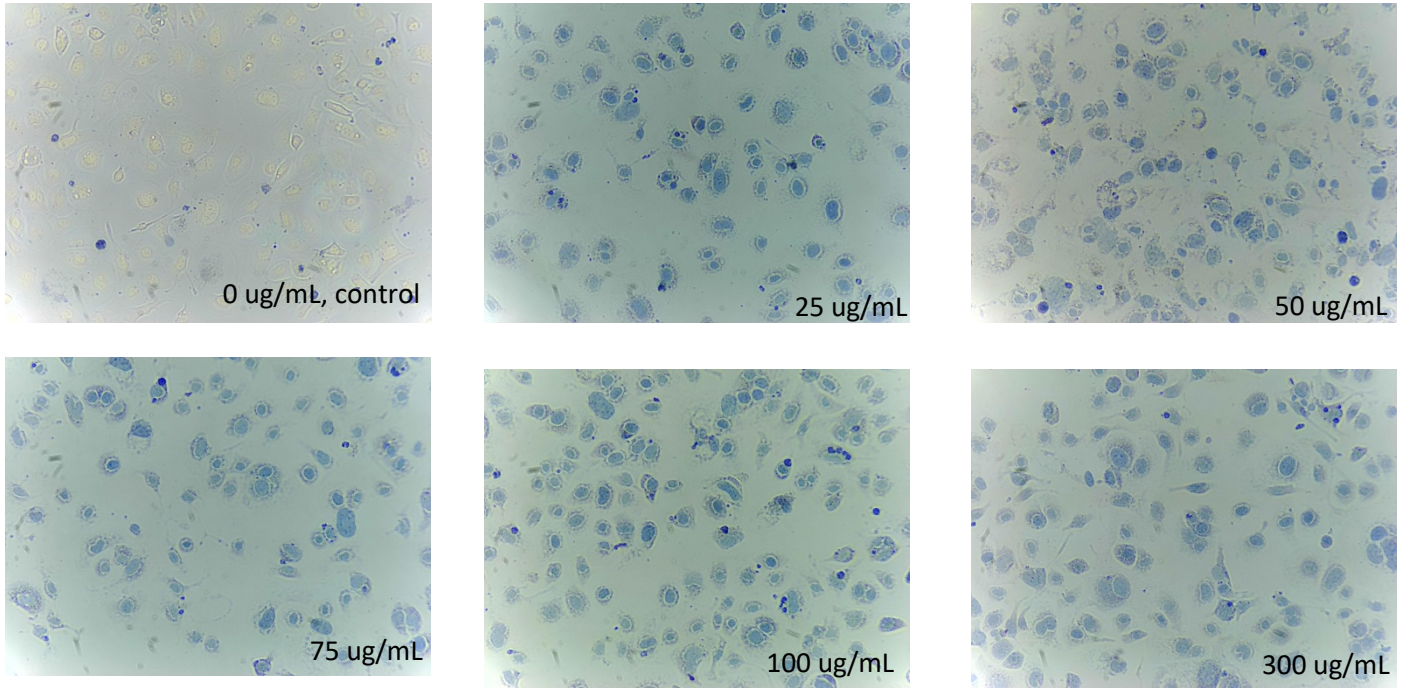


Figure 4.10 Images of HL-1 cells tested with different concentrations of digitonin and stained with trypan blue as visualised under the microscope at 40X magnification.

Following administration of different concentrations of digitonin ranging from 25ug/ml to 300ug/ml with untreated cells as the negative control (0ng/µl), the cells were stained with trypan blue to assess for cell permeabilisation.

The steps in methods chapter Section 2.7.2 were then conducted to isolate the mitochondrial fraction of the cells for subsequent GC-MS analysis. **Figure 4.11 (a)** shows the chromatogram from the non-transfected mitochondrial replicates. No signal for D-glutamate was detected above the noise from the mitochondrial extracts. As shown in **Figure 4.11 (b)**, L-glutamate is detected from both non-transfected and transfected whole cells and mitochondrial extracts. However, higher levels were detected from the whole cells indicating that the predominant glutamate pool appears to be cytosolic. Overall, D-glutamate was not detected in either whole cell or mitochondrial extracts.

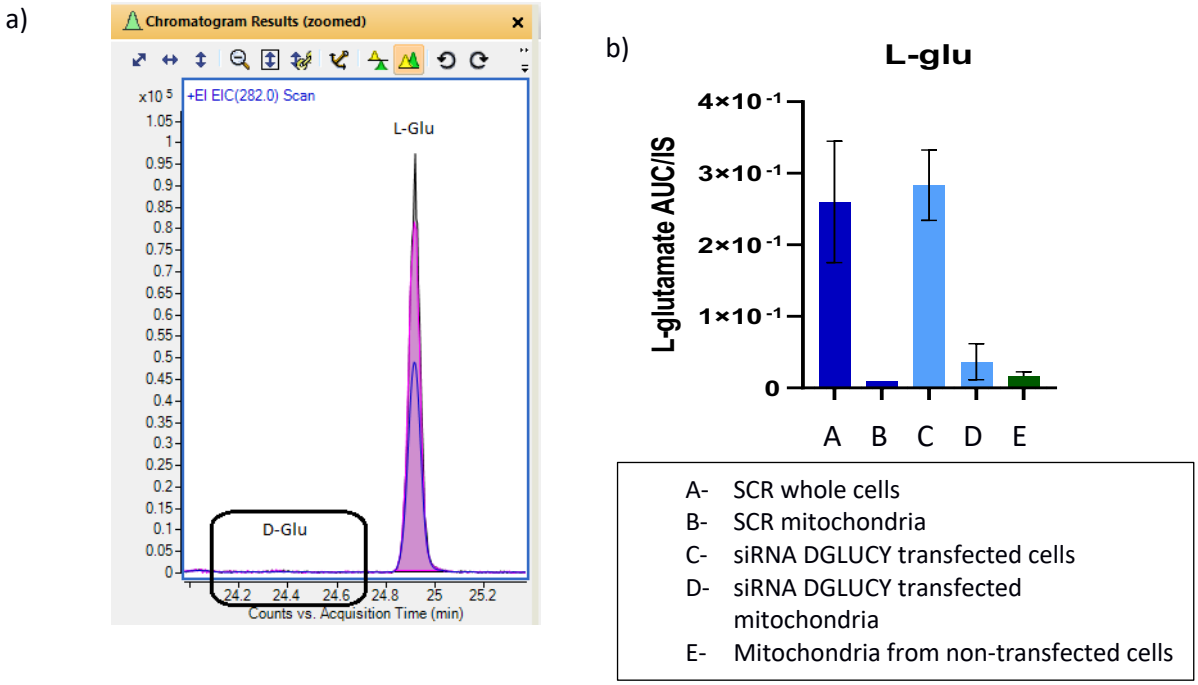


Figure 4.11 a) Chromatogram showing the peak areas of D- and L- glutamate isomers extracted from a pool of mitochondrial replicates that were non-transfected. b) Comparison of L-glutamate levels between whole cells and mitochondrial extracts

a) The x-axis represents the retention time of the D- or L-glutamate isomers, and the y-axis represents the ion counts (%). b) The x-axis represents the different HL-1 cell conditions: Dark blue bars- whole cells and mitochondrial extracts from SCR HL-1 cells, light blue- cell and mitochondrial extracts from cells transfected with DLGUCY siRNA, green- mitochondrial extracts from non-transfected cells. The y-axis represents the AUC L-glutamate levels detected for the different cell conditions. Abbreviations: AUC- area under the curve, L-glu- L-glutamate, mitos/MITO- mitochondrial. Image obtained courtesy of Jennie Roberts

4.2.2 Genetic determinants of fluoropyrimidine cardiotoxicity

4.2.2.1 Are variants recommended for testing before treatment with 5-FU associated with cardiovascular toxicity?

As previously discussed in Chapter 1 Section 1.7.2.1, *DPYD* variants are known to be associated with severe fluoropyrimidine toxicity, four of which are currently being recommended by the Clinical Pharmacogenetics Implementation Consortium (CPIC) to be screened before administering 5-FU or capecitabine-based chemotherapy. Using the QUASAR2 clinical trial dataset, which includes colorectal cancer patients treated with capecitabine, I tested the performance of the known *DPYD* risk variants in predicting cardiac toxicity events in this cohort. Results of the summary statistics of these variants are shown in **Table 4.1**. c.1129–5923C>G/HapB3 variant was significantly associated with increased risk of cardiac toxicity (OR= 4.28, 95% CI= 1.4-13, P=0.03). Other *DPYD* risk variants tested are rare with low allele frequencies (MAF<0.01); therefore, the analyses will likely be underpowered to identify any significant associations, reflected by their extreme effect sizes and wide confidence intervals.

Table 4.1 Association of *DPYD* risk variants with cardiac toxicity under the additive model

DPYD variant	RSID	Chr	Pos	Effect allele	MAF	OR	95% CI	P value
c.1129–5923C>G/HapB3*	rs56038477	1	98039419	T	0.02	**4.28**	**1.40-13.04**	**0.03**
c.2846A>T*	rs67376798	1	97547947	T	0.005	2366.27	$1.80x10^{-41}$ - $3.11x10^{47}$	0.43
c.1905+1G>A*	rs3918290	1	97915614	T	0.004	0.001	$1.03x10^{-35}$ - $1.02x10^{29}$	0.48
c.1679T>G*	rs55886062	1	97981343	C	0.0005	NA	NA	NA
Any CPIC variant						**2.51**	**0.88-7.16**	**0.13**
c.1601G>A	rs1801158	1	97981421	T	0.02	0.89	0.12-6.77	0.91

The four *DPYD* risk variants recommended for screening by CPIC (annotated with *) were tested for their association with cardiac toxicity in the QUASAR2 trial cohort using the 'method expected' model in SNPTEST. The ability to detect cardiac toxicity with any CPIC variant risk allele and another known DPYD risk variant c.1601G>A not currently used in screening is also being tested.

4.2.2.2 Meta-analysis of summary statistics from three colorectal cancer clinical trials investigating genetic determinants of cardiovascular toxicity following treatment with 5FU-containing regimens

A meta-analysis was conducted based on the GWAS summary statistics from the COIN, SCOT and QUASAR2 clinical trials. The trial design and cohort details are described in Chapter 2: Section 2.1 and summarised in **Table 4.2**.

Table 4.2 Three large colorectal cancer trials included in the meta-analysis of cardiotoxicity

Study	Sample size	Population	Treatment regimes	No. of patients with cardiac toxicity
QUASAR2	930	Stage 2/3 CRC	Capecitabine vs Bevacizumab + Capecitabine	29
COIN	1735	Advanced CRC (inoperable metastatic/locoregional measurable)	Arm A) Oxaliplatin + 5-FU/capecitabine Arm B) Oxaliplatin + 5-FU/capecitabine + cetuximab Arm C) Intermittent Oxaliplatin + 5-FU/capecitabine	46
SCOT	2939	High-risk stage 2/3 CRC	FOLFOX vs XELOX	39

Summary of the sample population and fluorouracil treatment received is described. Abbreviations: CRC- colorectal cancer, FOLFOX- IV 5-FU, folinic acid and oxaliplatin, XELOX- capecitabine and oxaliplatin

A total of 5604 patients with colorectal cancer treated with either 5-FU or capecitabine-based chemotherapy regimens were included in the meta-analysis. **Figure 4.12** shows the Manhattan plot generated from the meta-analysis. In addition to the chromosome 14 locus which was identified previously, I identified two new loci associated with cardiovascular toxicity. The first, rs141495755 on chromosome 7 maps to *TPK1* gene and was the top independent hit significantly associated with cardiac toxicity at GWAS significance ($P=9.64x10^{-9}$) with data available from all three study cohorts **(Table 4.3)**. A second SNP, rs117095544 reached GWAS significance ($P=1.37x10^{-8}$) with data from only two cohorts, QUASAR2 and COIN.

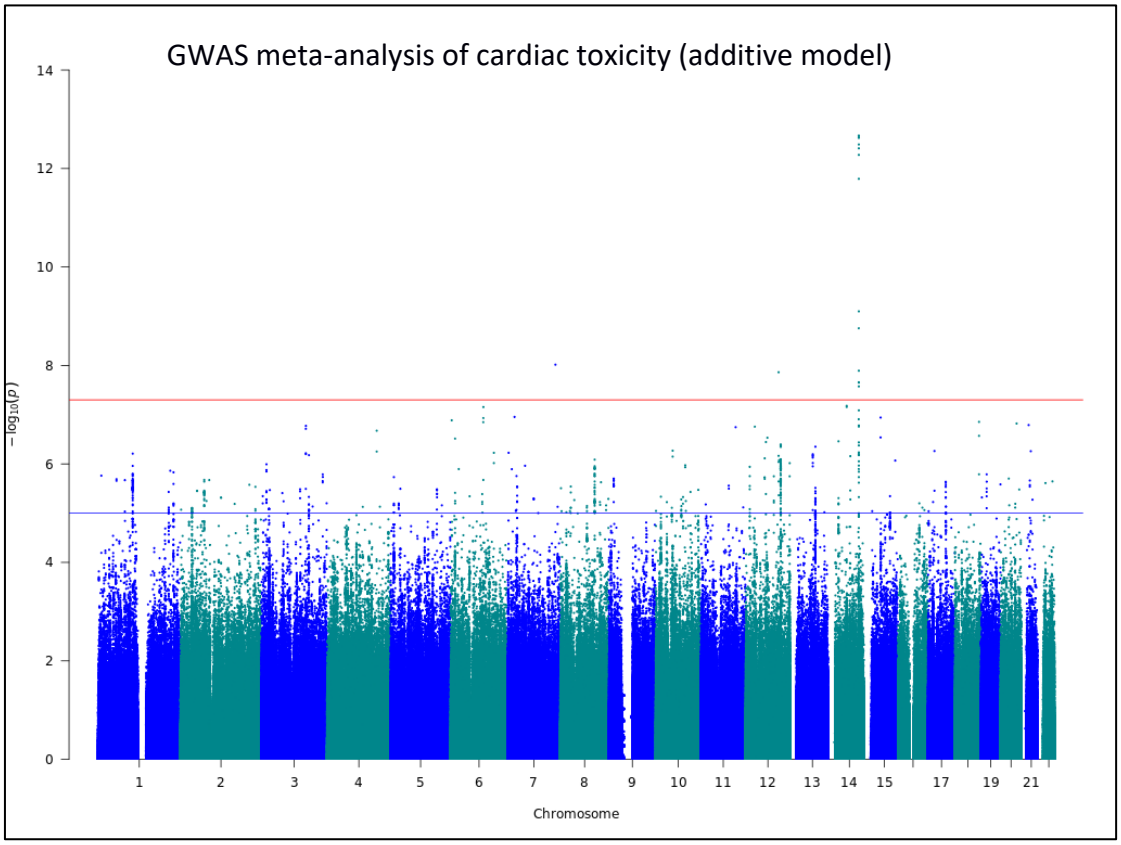


Figure 4.12 Manhattan plot showing results from the GWAS meta-analysis of cardiac toxicity from 3 trial cohorts (QUASAR2, COIN and SCOT)

GWAS outputs for the meta-analysis of cardiac toxicity under the additive genetic model are shown. The x-axis represents the SNPs from different chromosomes, and the y-axis represents the significance level of association at -log10 of the P value of each SNP with the presence of fluoropyrimidine induced cardiotoxicity. The red and blue lines represent the P value threshold of $<5x10^{-8}$ (GWAS significance level) and $<5x10^{-5}$. Abbreviations: GWAS -genome-wide association study

Table 4.3 Top SNPs identified from GWAS meta-analysis of fluoropyrimidine induced cardiac toxicity from 3 trial cohorts under the additive genetic model

rsID	Chr	Pos	ref A	Freq	OR	95% CI	P value	Consequences	Mapped Genes
rs141495755 ***	7	144498917	T	0.02	5.63	3.12-10.15	9.64E-09	intronic	TPK1
rs117095544 **	12	99233430	A	0.02	6.26	3.32-11.78	1.37E-08	intronic	ANKS1B

Conditional and joint ('COJO') analysis was used to identify top independent SNPs associated with cardiac toxicity that reached a P-value threshold of $<5x10^{-8}$ after filtering for MAF >1%, info score > 0.8, HWE $1x10^{-6}$, missingness 3% and P heterogeneity >0.05. refA refers to the allele associated with toxicity risk. *** present in all 3 trial cohorts, ** present in QUASAR2 and COIN. Abbreviations: Chr- chromosome, CI- confidence interval, Freq- risk allele frequency, OR- odds ratio, Pos- genomic position (build GRCh37)

Figure 4.13 shows the circos plot generated from chromatin interaction mapping for chromosome 7 that contains the top risk variant (chr7: 144498917) identified from the GWAS meta-analysis. One of the intra-chromosomal interactions identified between region 1: 7:144480001-144520000 (containing the genomic risk loci) and region 2: 7:144440001-144480000 (mapping associated genes) occurs within the left ventricle (FDR of the interaction: $2.6x10^{-12}$, Hi-C data from GSE87112 database).

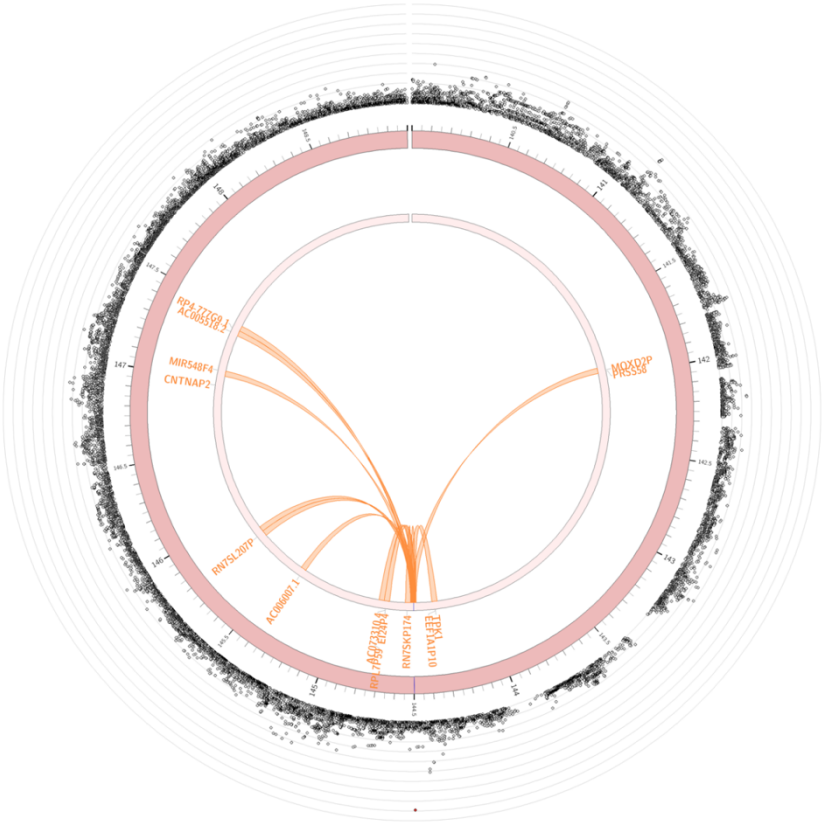


Figure 4.13 Chromatin interaction mapping for chromosome 7 based on GWAS meta-analysis of cardiac toxicity (additive model)

Build-in chromatin interaction mapping was performed using FUMA. The most outer layer of the circos plot is the Manhattan plot, where only SNPs with P<0.05 are displayed. Red dots are SNPS with r2>0.8 and grey dots are SNPS which are not in LD (r2 ≤0.2) with any independent significant SNPs in the locus. Genes that are mapped by chromatin interaction only are coloured in orange. The third layer is the chromosome ring which is the same as the second layer but without coordinates. Chromatin interaction links are coloured in orange.

I repeated the meta-analysis using a recessive genetic model. **Figure 4.14** shows the Manhattan plot generated from the meta-analysis. Nine variants reached GWAS significance ($P \leq 5x10^{-8}$) in their association with cardiac toxicity and were present in all three study cohorts. rs2911298A>C, the most significant SNP ($P= 7.35x10^{-10}$) is located in an intergenic region, nearest gene *CDH8* **(Table 4.4).**

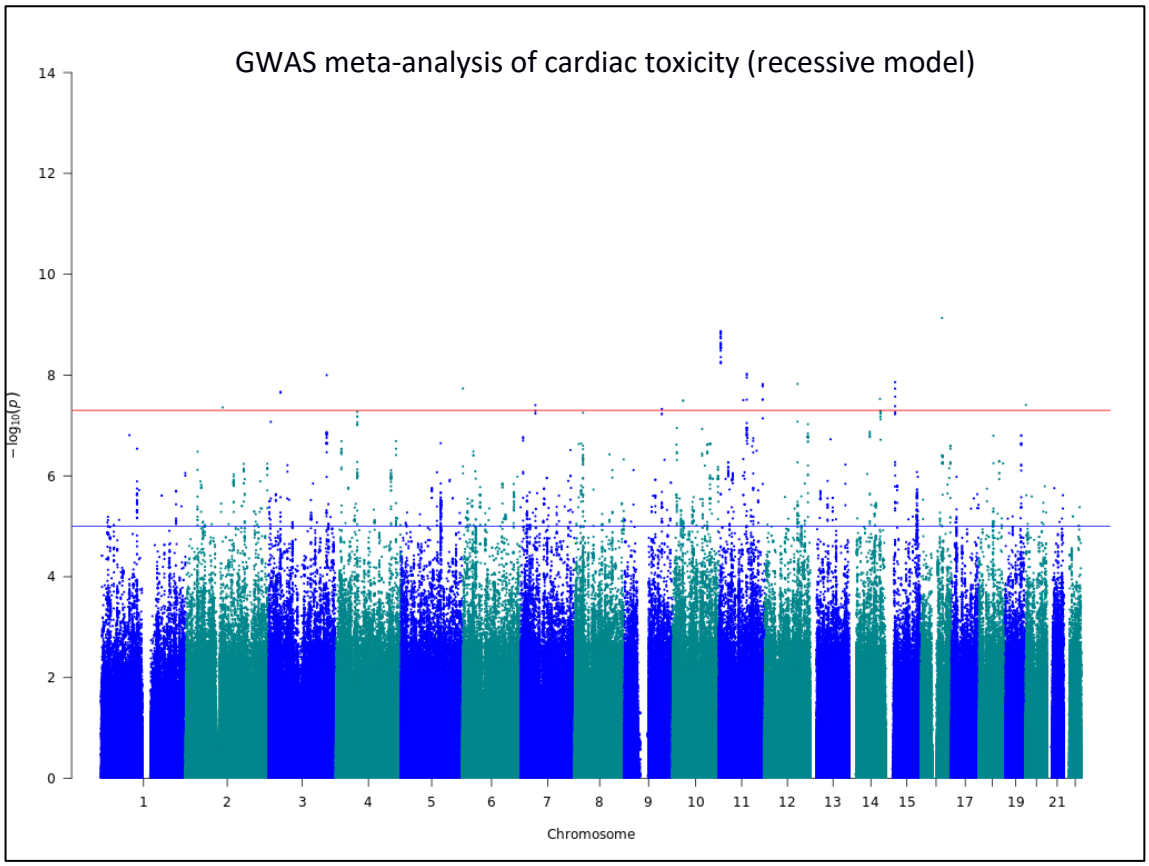


Figure 4.14 Manhattan plot showing results from the GWAS meta-analysis of cardiac toxicity from three trial cohorts (QUASAR2, COIN and SCOT)

GWAS outputs for the meta-analysis of cardiac toxicity under the recessive genetic model are shown. The x-axis represents the SNPs from different chromosomes, and the y-axis represents the significance level of association at -log10 of the P value of each SNP with the presence of fluoropyrimidine induced cardiotoxicity. The red and blue lines represent the P value threshold of $<5x10^{-8}$ (GWAS significance level) and $<5x10^{-5}$. Abbreviations: GWAS -genome-wide association study

Table 4.4 Top SNPs identified from GWAS meta-analysis of fluoropyrimidine induced cardiac toxicity from three trial cohorts under the recessive genetic model

rsID	Chr	Pos	refA	freq	OR	Lower CI	Upper CI	P value	Consequences	Mapped/Nearest gene
rs2911298	16	62623634	C	0.16	13.43	5.88	30.7	7.35E-10	intergenic	CDH8(dist=552895)
rs10838824	11	5933535	T	0.07	24.97	8.82	70.7	1.35E-09	intergenic	OR52E5(dist=10545),OR56A3(dist=35042)
rs55862501	11	82431596	T	0.11	11.22	4.92	25.62	9.41E-09	intergenic	MIR4300HG(dist=316641),FAM181B(dist=11450)
rs12793365	11	1.29E+08	T	0.12	7.56	3.75	15.24	1.50E-08	intergenic	LOC399975(dist=82416),BARX2(dist=10406)
rs56339115	12	96553313	A	0.11	10.03	4.51	22.27	1.50E-08	intergenic	LTA4H(dist=116015),ELK3(dist=34847)
rs72836734	6	1513316	A	0.06	50.45	12.87	197.77	1.85E-08	intergenic	FOXF2(dist=117478),FOXCUT(dist=92450)
rs61893917	11	1.72E+08	C	0.09	17.78	6.42	49.28	3.13E-08	intergenic	CLPB(dist=115585),LINC01537(dist=20387)
rs9639876	7	43512183	A	0.05	22.86	7.48	69.82	3.94E-08	intronic	HECW1
rs55693048	9	1.1E+08	G	0.13	8.1	3.82	17.16	4.71E-08	intergenic	LINC01505(dist=106078),ZNF462(dist=77219)

Conditional and joint ('COJO') analysis was used to identify top independent SNPs associated with cardiac toxicity that reached a P-value threshold of $<5x10^{-8}$ after filtering for MAF >1%, info score > 0.8, HWE $1x10^{-6}$, missingness 3% and P heterogeneity >0.05. refA refers to the allele associated with toxicity risk. All SNPS are present in all three trial cohorts. Abbreviations: Chr -chromosome, CI- 95% confidence interval, freq- risk allele frequency, OR- odds ratio, Pos- genomic position (build GRCh37)

4.3 Discussion

4.3.1 Determining the functional role of *DGLUCY* in the presence of 5-FU

Treatment with certain doses of 5-FU (10uM and 100uM when treated once and 5uM when treated twice) reduced cell viability compared to untreated cells, but no decrease in cell viability was observed with increasing 5-FU concentration. Successful quantification of target genes *DGLUCY* and *RPS6KA5* was only possible from cells treated with 0.1µm of 5-FU. Target gene expression in these cells was no different to that in untreated cells. It appears that treatment with 5-FU does not affect expression of these genes though perhaps there is a decrease in cell viability at higher concentrations as we could not detect housekeeping or target gene expression at higher concentrations. It is, however, strange that the MTS assay did not show lower cell viability with higher 5-FU concentrations among the treated cells.

A significant limitation of the work aiming to determine if D-glutamate was detectable in cells is the lack of a positive tissue or cell sample control. We confirmed the column's ability to distinguish between synthetic L and D glutamate, but we could not detect D-glutamate in the experiments described. Levels were possibly below the detection limits, but we also note that Ariyoshi et al., which was the first study that successfully detected high levels of D-glutamate in 9030617O03Rik-deficient mouse hearts so far, have used 2D-HPLC to detect the D-enantiomers, which relies on stereospecific derivatisation and non-chemical-based detection (absorbance in visible spectrum)[187]. A high L:D glutamate ratio was observed in the study from mouse heart tissue; thus, using mouse cardiomyocyte cell lines in our experiments may have rendered it difficult to reach the concentration level required for detecting D-glutamate[187]. There is also no direct chemical evidence for the existence of D-enantiomers in mammalian cells yet, indicating more functional work will need to be derived to achieve the

optimal detection of D-enantiomers in 9030617O03Rik-deficient or wild-type mammalian cells.

Due to the limited existing literature assessing the relationship between *DGLUCY* and cardiac function or toxicity, most experimental designs were limited to methods based on two publications[187,221]. However, the inability to detect the D-enantiomer from our cardiomyocyte cell line model indicates that heart tissue may be required for future functional experiments assessing the effects of *DGLUCY* on cardiac function.

4.3.2 Identifying genetic variants predisposing towards fluoropyrimidine cardiotoxicity

One of the *DPYD* risk variants currently used in screening tests for fluoropyrimidine toxicity, c.1129–5923C>G/HapB3, was significantly associated with cardiac toxicity from the QUASAR2 dataset. To date, there are limited studies assessing the effectiveness of these *DPYD* risk variants in predicting cardiac toxicity, which can be severe and life-threatening. Case reports of patients who developed ischaemic heart disease and cardiomyopathy post 5-FU treatment found they carried germline *DPYD* mutation p453L (1358C>T)[24], *DPYD* c.85T>C mutation and *TYMS* 5'TSER genotype 2R/2R respectively[226]. Establishing that c.1129–5923C>G can help identify patients at risk of cardiac events from this analysis is an important finding that reaffirms the validity of screening for this variant in practice.

There is currently no established genetic biomarker predictive of fluoropyrimidine-induced cardiotoxicity. My group previously identified a common SNP on chromosome 14 as a novel biomarker of 5-FU-induced cardiovascular toxicity. By conducting a GWAS meta-analysis of three cohorts of CRC patients treated with 5-FU/capecitabine based therapies, I have identified two additional independent markers, rs141495755 and rs117095544. Both SNPs are

rare (MAFs 0.02), but rs141495755 has data from three cohorts, all showing consistent effects. The role of rs141495755 in heart function or cardiac toxicity is still unknown. rs141495755 maps to *TPK1,* which encodes thiamine phosphokinase 1, a protein which catalyses the conversion of thiamine to thiamine pyrophosphate (TPP), which acts as a cofactor for the enzymes involved in the glycolytic and energy production pathways[227]. TPP is the active form of thiamine and is essential in the energy production of several organ systems including cardiac function[228]. Insufficiency of TPP has been associated with acute heart failure[229]. A study by Polat et al. assessed the effect of thiamine and TPP on cardiotoxicity induced by doxorubicin in rats and found that the cardiotoxic effects were caused by the lack of thiamine phosphokinase enzyme leading to low TPP levels[228]. Another study by Coskun et al. found that TPP prevented cisplatin chemotherapy-induced oxidative stress and cardiotoxicity in rats[230]. The findings of these studies suggest TPP has a protective effect against cardiac injury induced by chemotherapy; thus, individuals with mutations in *TPK1* are likely to have problems producing sufficient amounts of TPP, causing them to be at high risk of cardiac events. Phenome-wide association studies (PheWAS) have found associations between rs141495755 with cardiac-related phenotypes, including coronary artery bypass grafting with P= 7.4 $x10^{-4}$ and pulse rate with P= 8.7$x10^{-4}$ from the FinnGen and UK Biobank datasets respectively[214,231,232]. Although further fine mapping of the SNP showed that it is less likely to influence regulatory elements of TPK1 with a high regulome score of 6, Hi-C data shows significant chromatin interaction with other genomic regions within chromosome 7 in the left ventricle, highlighting other potential candidate genes of interest.

I also identified nine independent SNPs associated with cardiac toxicity with GWAS significance under the recessive model that are present in all three study cohorts. PheWAS found that SNPs in chromosome 11 such as rs10838824 have been linked with valvular operations, heart disease and atrioventricular block, while rs61893917 have been linked with

atrial fibrillation and flutter from the FinnGen database[214,231]. rs55693048 in chromosome 9 has also been associated with high cholesterol, a cardiovascular disease risk factor from UK BioBank data[233]. These cross-phenotype associations indicate that these variants could have the ability to predict intermediate and direct phenotypes related to risk of cardiac toxicity.

4.3.3 Future directions

DGLUCY variants rs4900071 and rs4900072 showed promising potential in their predictive ability of fluoropyrimidine induced cardiotoxicity. Further functional experiments may require mouse heart tissue instead of cell lines to determine the interaction between 5-FU and *DGLUCY* gene and protein expression. In addition to the *DGLUCY* variants, the *DPYD* risk variant c.1129–5923C>G/HapB3 and the novel variants I identified by meta-analysis of cardiac toxicity from three large study cohorts can help contribute to future screening panels to identify patients at high risk of fluoropyrimidine-induced cardiac toxicity. A pilot study to investigate the sensitivity and specificity of the SNPs in clinical practice would be the next step required. Successful validity of their predictive ability in practice will have important implications in preventing a potentially life-threatening complication from their treatment.

CHAPTER 5: IDENTIFICATION OF GENETIC FACTORS PREDISPOSING TO IMMUNE CHECKPOINT INHIBITOR TOXICITY

5.1 Introduction

Immune checkpoint inhibitors (ICI) have established themselves as a key therapeutics against cancer. ICI are highly efficacious in treating certain cancers such as melanoma, with high response rates of over 50% being recorded[31]. However, ~ 50% of patients receiving doublet ICI therapy or ~10% receiving single agent ICI therapy experience grade 3 or higher immune-related adverse events (irAE)[234]. The management of irAE remains a challenge in clinical practice as the spectrum of irAE is broad and can affect multiple organ systems[28]. The time to onset of different irAE can also vary according to the type of irAE and ICI treatment regime received[28].

There are currently no predictive biomarkers that can detect patients at risk of serious irAE. Our host genome has been strongly hypothesized to be a key determinant of irAE given its strong influence on immune homeostasis, which could impact the immune response to ICI therapies[52]. The role of germline genetics in predicting irAE is supported by evidence of the association between germline genetic markers and expression of immuno-modulatory cells/proteins (including immune checkpoints) from the cancer immunity cycle (Chapter 1, Figure 1.4). Thousands of disease risk variants identified from GWAS of various autoimmune disease traits, which closely resemble the clinical presentation of irAE, further support this hypothesis. Several studies have since assessed the correlation of genetic factors with irAE development (Chapter 1, Table 1.7). Most studies are small to moderate-sized candidate gene studies evaluating a selected panel of SNPs mapping to the MHC loci (HLA genes) or immune checkpoint pathway such as PD-1/CTLA4. A more recent large study of >1000 patients treated with ICI identified a variant mapped to IL-7, which was associated with increased risk of any grade irAE at a GWAS significant level (HR 2.0 (1.6-2.5), $p=3.8 \times 10^{-9}$)[82].

As previous studies have shown a positive correlation between ICI treatment response and toxicity[137,138], it would be interesting to assess whether there are shared susceptibility loci that could contribute to both treatment outcomes. A study of 94 non-small cell lung cancer (NSCLC) patients found distinct sets of SNPs from different biological pathways correlated with both treatment outcomes[122].

Given the potentially strong role of germline genetic factors as determinants of irAE, there is a need for more large population analyses such as GWAS to achieve sufficient power to detect novel SNPs associated with irAE caused by ICI. Large sample sizes are often required to achieve sufficient power to identify common SNPs with clinically relevant effect sizes. Assuming that the rate of serious irAE is 10%, an estimated sample size of >500 patients is required to achieve a power of >80% to identify common variants (minor allele frequency (MAF) >30%) with effect sizes (Odds ratio) >1.8 and rare variants (MAF <10%) with OR >2.0. Previous studies that have assessed for genetic markers predictive of ≥grade 3 irAE (eg. Marschner et al with 167 patients [129] and Groha et al with 1751 patients [82]) had sufficient power (>80%) to identify common variants with OR>3. Using large GWAS coupled with polygenic risk score profiling will help develop a potential candidate gene panel that can be used as a predictive risk model of toxicity in clinical practice.

5.1.1 Chapter aims

This chapter had the following aims:

1. Identify germline SNPs that are predictive of ≥grade 3 irAE (both organ-specific toxicities and toxicities of any type) and validate the top SNPs in a replication cohort
2. Generate polygenic risk scores for detecting irAE based on known loci associated with autoimmune disease and immune-modulatory traits
3. Identify rare damaging or disruptive variants predictive of ≥grade 3 irAE

5.2 Results

5.2.1 Recruitment to ICI GENETICS

From April 2019 to October 2022, I coordinated the recruitment of 775 participants from 40 sites for the ICI GENETICS study. The eligibility criteria for this study includes patients who are treated with ICI therapy, and recruited either prospectively (prior to or during the first 2 cycles of their treatment) or retrospectively (patients who have experienced a ≥grade 3 irAE). Other inclusion criteria include ≥age 16 and agreement to provide a sample of blood or saliva. I opened the study at most of these sites and supported the sites by sending blood sampling resources, checking the toxicity data collected and answering study queries. **Figure 5.1** demonstrates the number of patients recruited from each site and **Table 5.1** lists their baseline demographics and first-line ICI treatment received.

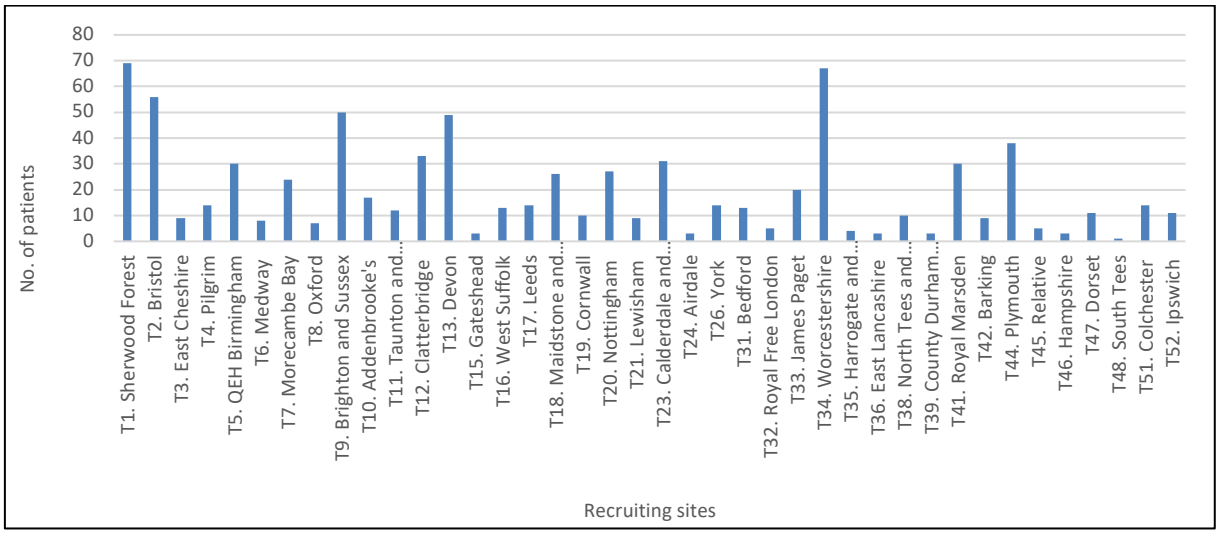


Figure 5.1 Participants recruited to the ICI GENETICS Study (IRAS 237779) up till October 2022
The x-axis shows the name of the recruiting hospital site and the y-axis represents the number of participants recruited in each site.

Table 5.1 Baseline characteristics of patients recruited to the ICI GENETICS study

Demographics	**No. of patients (N=775)**
Mean Age (range)	67 (28-91)
Sex	
Male	440
Female	335
Ethnic origin	
White	760
Black/Black British	6
Asian/Asian British	4
Chinese/Other ethnic	1
NA	4
Cancer diagnosis	
Lung	366
Skin/Melanoma	220
Kidney	61
Bladder/Ureter	35
Head and neck	30
Upper GI	18
Breast	15
Colorectal	15
Unknown primary	4
Prostate	3
Gynaecological	2
Hepatobiliary	2
Haematological	1
NA	3
ICI regime	
Anti-PD1	534
Anti-PDL1	127
Anti-CTLA4+anti-PD1	105
Anti-CTLA4	5
NA	4

Cancer diagnoses and ICI agents received as first-line treatment by 775 patients recruited to the study until October 2022. Abbreviations: GI- gastrointestinal, ICI- immune checkpoint inhibitor, NA- data not available

5.2.2 Genome-wide association studies of ICI toxicity

5.2.2.1 Sample selection and initial genotype quality control for GWAS Discovery Phase

566 patients with DNA extracted from blood available were initially sourced from a collaboration between six study groups (ICI GENETICS, PAIR, MOLGEN/HYST, PRIMM, MEDALLION and Oxford) **(Figure 5.2)**[235]. These samples were sent to genotyping on whole genome SNP tagging arrays as detailed in methods chapter Section 2.13.

A large number of samples included in the first batch for genotyping had a low call rate (N=159 with call rate <97%). The technical controls (HapMap DNAs) included by the Genotyping facility passed the 97% cut off but had lower call rates than seen in the other batches (0.98 versus 0.99 for other genotyping batches). The genotyping facility were unable to help us resolve the issue with this batch of samples and with no clear explanation for their low performance we had to exclude the low call rate samples resulting in 407 patient samples available for analysis.

After merging with the HapMap population and performing PCA, 14 patients were excluded (six due to relatedness, three duplicate samples and five as PCA outliers). To understand if there was an issue with sample quality, we resent 24 failed samples in a further batch for genotyping. 17/24 passed call rate ≥97% this time. In addition to these 17 patients, a second batch of 253 samples (229 patients from Oxford and 24 patients from ICI GENETICS) with genotyping data were included in the analysis. The same quality control (QC) steps were applied to the merged cohort and after phasing for haplotype estimation, a total of 451,117 variants and 654 patients remained. Following imputation and data cleaning, ~87 million variants and 599 patients with adequate toxicity data were available for analysis. The filtering steps described above are outlined in **Figure 5.2**.

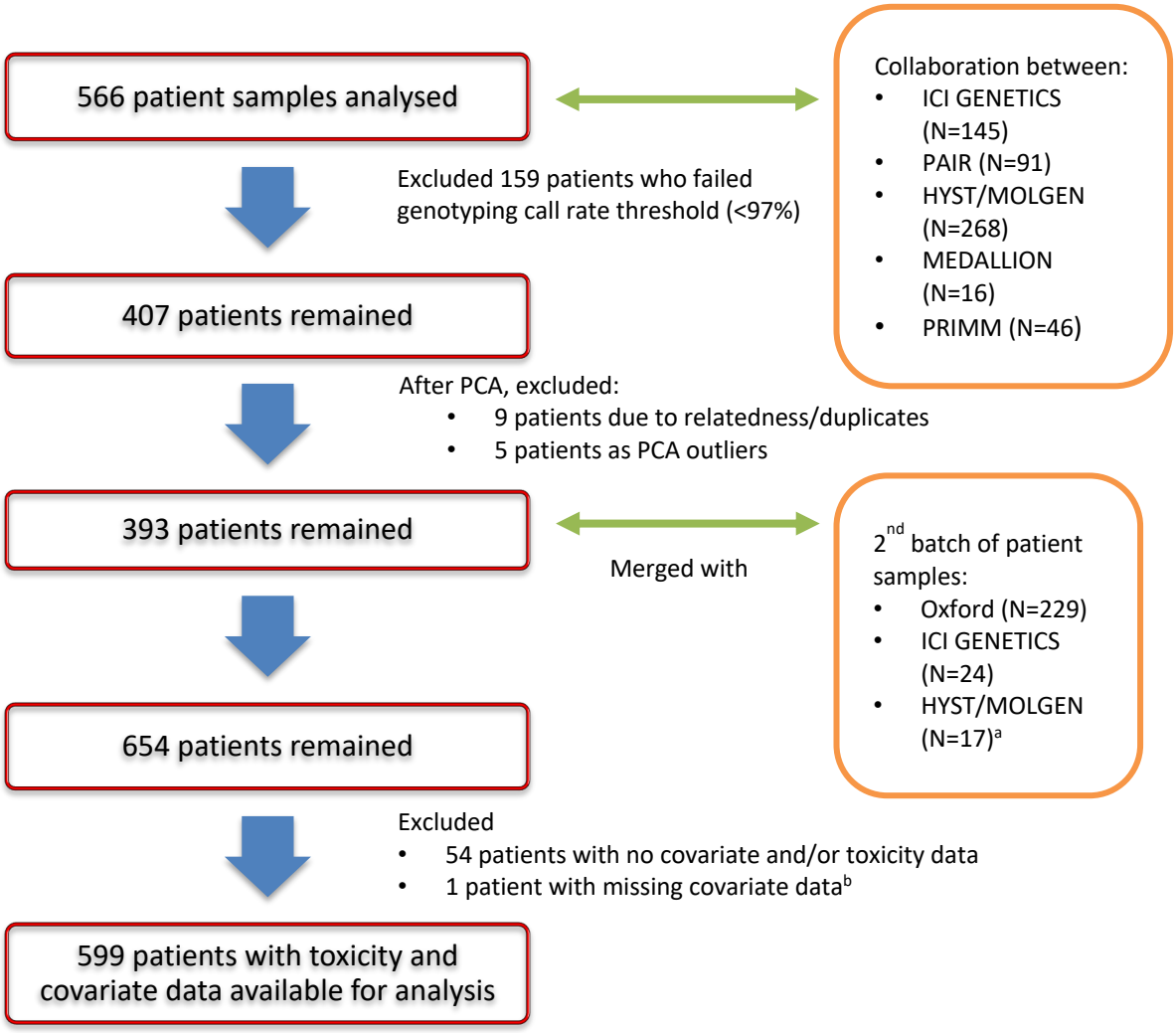


Figure 5.2 Consort flow diagram of outputs from the QC steps applied to the genotyping data of samples included in the GWAS analysis of irAE.

UK study groups collaborated to assemble a joint dataset of patients treated with ICI where toxicity data and DNA samples for genotyping were available. The QC filtering steps applied to exclude SNPs/samples from the dataset include: >3% missingness per variant and sample, Pihat >0.05 in pairwise sample comparisons indicating relatedness, +/-6 SD from the mean of PCs 1-10 suggesting outlier samples with divergent ancestry and rare variants with MAF of <1%. Abbreviations: ICI- immune checkpoint inhibitor, irAE- immune-related adverse event, QC- quality control, SD- standard deviation. [a]Samples that had passed the call rate threshold on repeat genotyping [b]No data on the presence of concurrent therapy were available

Patient characteristics and clinical information for the patients included in the GWAS analysis are provided in **Table 5.2**. Melanoma was the predominant cancer diagnosis (total n=430/599 (72%)), followed by lung and renal malignancies. The majority of patients with grade 3-4 events received treatment with a combination ICI (anti-CTLA4 and anti-PD1) regime (n=130/186 (70%) compared to single agent ICI therapy (n=45/343 (13%) (p<0.00001 by chi-

squared test). In both cohorts of patients with grade 0-2 and grade 3-4 toxicities, most did not receive any additional concurrent anti-cancer therapy with their ICI treatment.

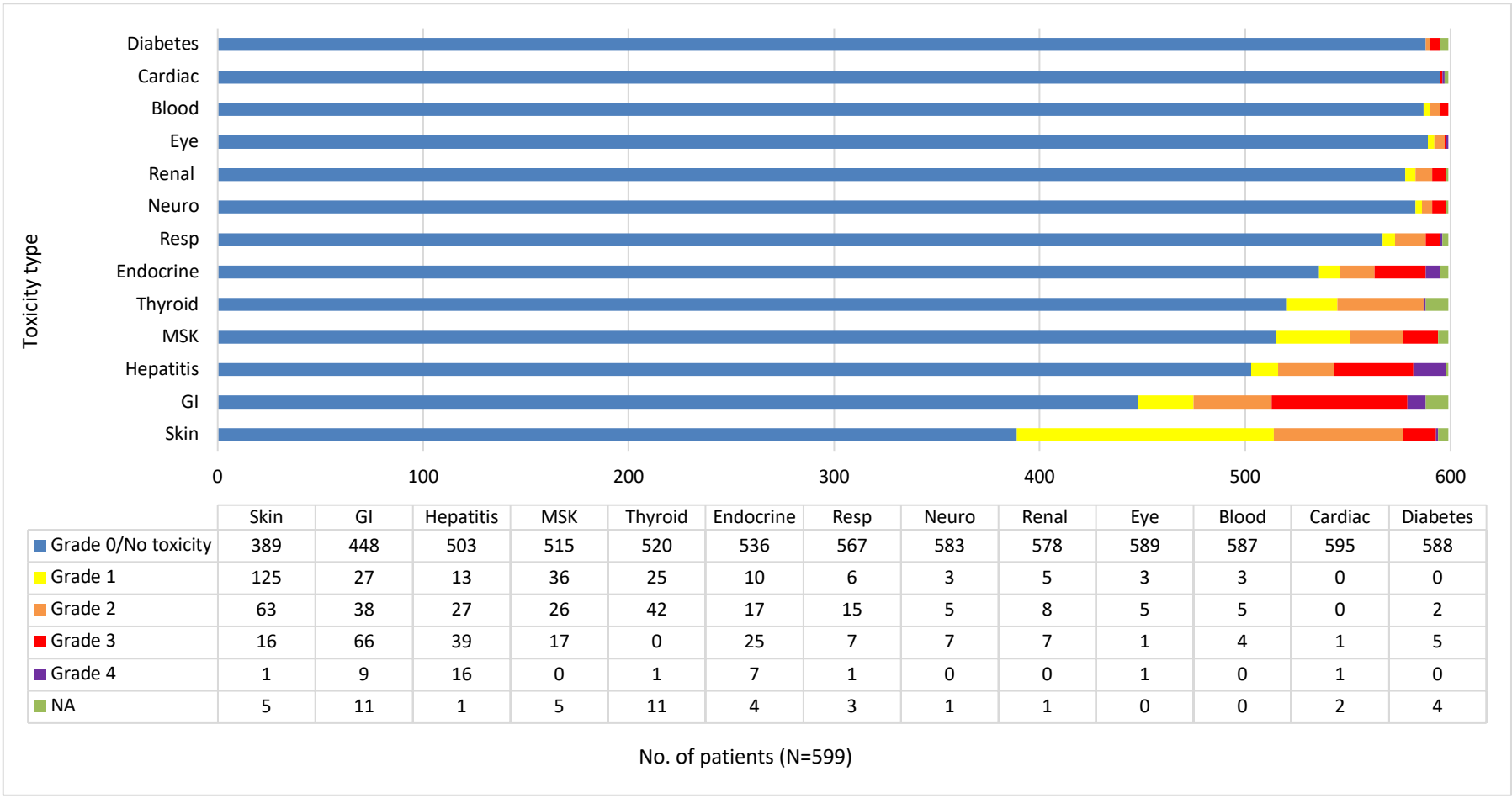
Table 5.2 Baseline characteristics of patients included in the GWAS of irAE who received ICI treatment

Demographics Total N=599	**No. of patients with grade 0-2 irAE, N=413 (69%)**	**No. of patients with grade 3-4 irAE, N=186 (31%)**
Sex		
Female	189 (46%)	76 (41%)
Male	224 (54%)	110 (59%)
Mean age (range)	66 (23-96)	61 (21-88)
Cancer type		
Skin/Melanoma	273 (66%)	157 (84%)
Lung/thoracic	96 (23%)	17 (9%)
Kidney	34 (8%)	12 (6%)
Head and neck	4 (1%)	0 (0%)
Bladder	2 (0.5%)	0 (0%)
Other	3 (0.7%)	0 (0%)
Unknown	1 (0.002%)	0 (0%)
ICI agent		
Anti-CTLA-4 + anti-PD1	95 (23%)	130 (70%)
Anti-PD1	268 (65%)	42 (23%)
Anti-PDL1	19 (5%)	1 (0.5%)
Anti-CTLA-4	11 (3%)	2 (1.1%)
Other (please specify)	1 (0.2%)	0 (0%)
Unknown	19 (5%)	11 (6%)
Concurrent Therapy		
Yes	17 (4%)	5 (3%)
No	395 (96%)	181 (97%)
Unknown	1 (0.2%)	0 (0%)

The total cohort of patients was gathered from six UK study groups and had data collected on the core fields required for analysis including: age, sex, type of irAE, maximum CTCAE grade, ICI treatment received, cancer diagnosis, and presence of any concurrent anti-cancer treatment. Abbreviations: ICI- immune checkpoint inhibitor, irAE- immune-related adverse event.

5.2.2.2 Toxicity profile of GWAS cohort

186/599 (31%) of patients in the discovery cohort developed ≥grade 3 irAE of any type, while 414/599 (69%) patients had grade 0-2 irAE. However, these frequencies may not be a true reflection of the real world incidence of irAE due to ascertainment bias from patients recruited retrospectively from ICI GENETICS who are already know to have a ≥grade 3 irAE. The number of patients in the genome-wide analysis experiencing different types of irAE according to grading is shown in **Figure 5.3**. Gastrointestinal toxicities formed the highest proportion of grade 3-4 events (n=75/186, 40%), followed by hepatitis and endocrine toxicities (n=55/186, 30% and n=32/186, 17% respectively. More than half of patients (≥54%) who developed hepatitis, gastrointestinal or endocrine irAEs were reported as ≥grade 3 events. The most common types of irAE recorded of any grade were skin toxicities (n=204/599, 34%), followed by gastrointestinal and hepatitis toxicity (n=140/599, 23% and n=95/599, 16% respectively).



	Skin	GI	Hepatitis	MSK	Thyroid	Endocrine	Resp	Neuro	Renal	Eye	Blood	Cardiac	Diabetes
Grade 0/No toxicity	389	448	503	515	520	536	567	583	578	589	587	595	588
Grade 1	125	27	13	36	25	10	6	3	5	3	3	0	0
Grade 2	63	38	27	26	42	17	15	5	8	5	5	0	2
Grade 3	16	66	39	17	0	25	7	7	7	1	4	1	5
Grade 4	1	9	16	0	1	7	1	0	0	1	0	1	0
NA	5	11	1	5	11	4	3	1	1	0	0	2	4

No. of patients (N=599)

Figure 5.3 Number of patients with CTCAE graded irAE included in the GWAS analysis.

The x-axis represents the number of patients who had received ICI, and the y-axis represents the organ-specific irAE experienced. The CTCAE grading is indicated with different colours; blue: grade 0/no toxicity, yellow: grade 1, green: grade 2, red: grade 3, purple: grade 4, grey: NA indicates that the toxicity is present but grading data was unavailable. Endocrine cases also included thyroid or hyperglycaemia (or diabetes) toxicity cases. Respiratory cases consisted of pneumonitis and cardiac irAE included myocarditis cases. Abbreviations: CTCAE- common terminology criteria for adverse events, GI- gastrointestinal, irAE- immune-related adverse events, MSK- musculoskeletal, Neuro- neurological, Resp- respiratory

5.2.2.3 Power calculation

When performing the GWAS of different toxicity variables, the sample sizes of cases to controls will vary according to the type of toxicity analysed. Hence for an estimated sample size of 50 cases and 500 controls, the analysis is powered at >80% to detect common variants (MAF 30%) variants with an OR of ≥1.9 and to detect rare variants (MAF<10%) at OR 2.2. Analyses with fewer cases (20 cases and 530 controls) will be powered at >80% to detect common variants with OR 2.5 and rare variants with OR 2.9.

5.2.2.4 Association analysis of serious irAEs based on additive genetic models

Association testing using the frequentist approach to identify genetic variants linked with any serious irAE was performed based on the additive model and adjusted for co-variates age, sex, presence of concurrent cancer treatment and treatment with single or combination ICI therapy. As described in Chapter 2: section 2.2.2, the "serious toxicity" variable consisted of any ≥grade 3 irAE (including skin, gastrointestinal (GI), hepatobiliary, endocrine, thyroid, neurological, renal, musculoskeletal, blood and ocular toxicities), ≥grade 2 pneumonitis and any grade myocarditis, whereas the control groups did not experience these listed events. **Figure 5.4** shows the Manhattan plot of the results from GWAS of serious irAE. No SNPs reached genome-wide significance ($P \leq 5x10^{-8}$) in the analysis. **Table 5.3** details the top eight independent SNPs associated with serious irAEs that reached a P-value threshold of $P<5x10^{-6}$. rs71399956(insTTG) is the most significantly associated variant with an OR 2.54, 95% CI 1.79-3.61, with its P-value close to GWAS significance ($P=8.91x10^{-8}$). The SNP maps to the *GPR158* gene. The SNP with the next lowest P-value is rs35429660. This marker has had the lowest RegulomeDB score of the set of suggestive SNPs - 3a, indicating its likely to have a strong role in gene regulation[236].

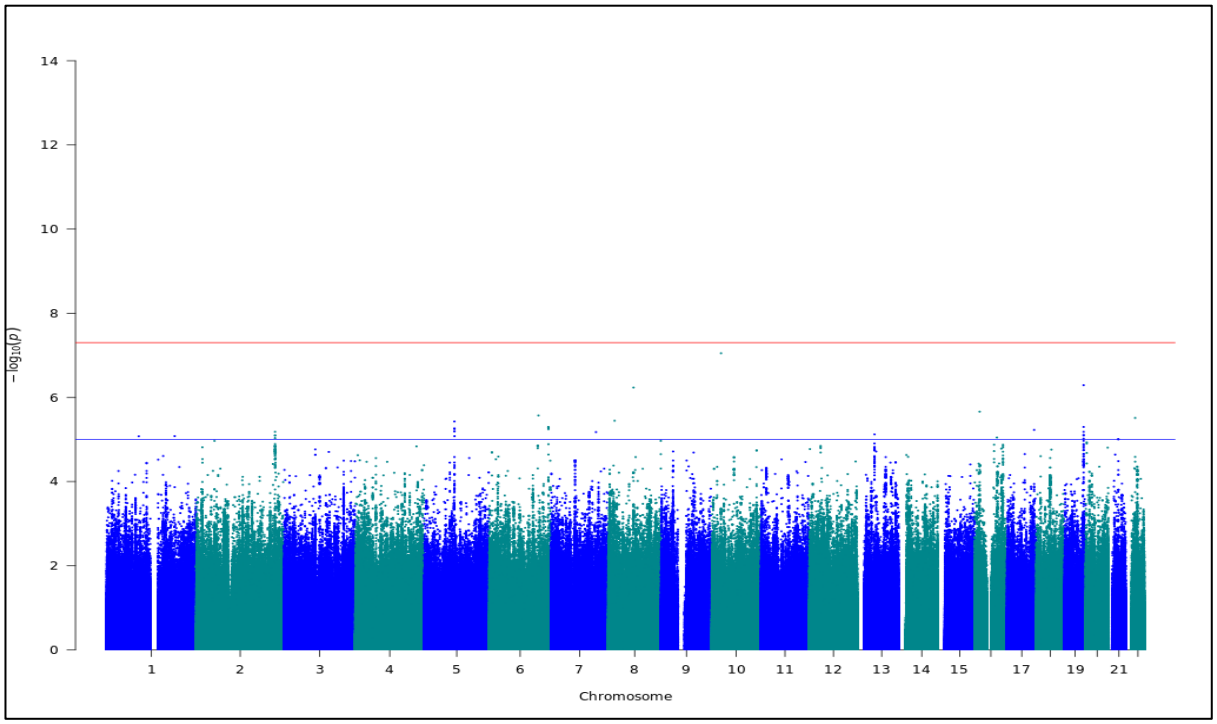


Figure 5.4 Manhattan plot from the genome-wide association analysis identifying variants predictive of serious irAE of any type.

The x-axis represents the SNPs from different chromosomes and the y-axis represents the significance level of association at -log10 of the P value for each SNP with any ≥grade 3 irAE. The red and blue lines represent the P value threshold of $<5x10^{-8}$ (GWAS significance level) and $<5x10^{-5}$. Abbreviations: irAE- immune-related adverse event

Table 5.3 Top independent SNPs identified from GWAS of serious irAE under the frequentist additive genetic model

Toxicity	rsID	Chr	Pos	Risk allele	Freq	INFO	OR	Lower CI	Upper CI	P value	Consequence	Mapped /nearest genes	Regulome DB
Any ≥grade 3 irAE	rs71399956	10	25497487	TTG	0.31	0.75	2.54	1.79	3.61	8.91E-08	intronic	GPR158	NA
	rs35429660	19	53088882	C	0.65	0.93	2.21	1.60	3.05	5.13E-07	UTR3	ZNF701	3a
	rs2472135	8	70006121	A	0.45	0.87	2.10	1.56	2.83	5.83E-07	ncRNA_intronic	LINC01592	7
	rs7197466	16	13993158	T	0.69	1	2.08	1.52	2.85	2.19E-06	intergenic	SHISA9(dist=658885), ERCC4(dist=20856)	6
	rs4527728	6	136358054	T	0.66	0.87	2.21	1.57	3.11	2.68E-06	intronic	PDE7B	5
	rs1389216733	22	25953500	CA	0.93	0.81	6.34	2.57	15.65	3.07E-06	intergenic	CRYBB2P1(dist=95855), GRK3(dist=7360)	NA
	rs397892660	8	17486594	CTT	0.81	0.96	2.48	1.65	3.71	3.59E-06	ncRNA_intronic	PDGFRL	NA
	rs6864863	5	83978011	A	0.63	0.92	2.10	1.52	2.91	3.73E-06	intergenic	EDIL3(dist=297326), NBPF22P(dist=1600251)	7

Conditional and joint ('COJO') analysis was used to identify the top independent SNPs associated with serious irAE that reached a P-value of $<5x10^{-6}$ after filtering for MAF>1% in cases and controls, and adjusting for covariates age, sex, single/doublet ICI therapy and presence of concurrent anti-cancer treatment. The risk allele refers to the allele associated with toxicity. Variant annotation was performed using gene annotation software tool ANNOVAR. RegulomeDB scores (from 1a to 7) were obtained from RegulomeDBv1.1 using FUMAv1.4.1; the lowest score represents SNPs with the most biological evidence as a regulatory element [236][84]. Abbreviations: Chr- chromosome, CI- confidence interval (95%), Freq- frequency of risk allele, INFO- info score, irAE- immune-related adverse event, MAF- minor allele frequency, NA- not available, ncRNA- non-coding RNA, OR- odds ratio, Pos- genomic position (build GRCh37), SNP- single nucleotide polymorphism, UTR- untranslated region

Curated gene sets and gene ontology terms from MSigDB were tested using MAGMA gene set analysis. **Table 5.4** shows the top significant gene ontology sets include dicarboxylic acid catabolism (OR=2.26, P=8.83x10-5, N=13), genes upregulated in liver tissue upon HNF1A knockout (OR=1.30, P=1x10^{-4}, N=130), and positive regulation of homotypic cell-cell adhesion (OR=3.30, P=2x10^{-4}, N=6)[172,211].

Table 5.4 Top 10 gene sets identified from the GWAS of any serious irAE

Gene Sets	N genes	Beta	SE	P value	P bon
GO BP: dicarboxylic acid catabolic process	13	0.817	0.218	8.83E-05	1
Curated gene sets: Servitja liver hnf1a targets up	130	0.260	0.072	1.49E-04	1
GO BP: positive regulation of homotypic cell-cell adhesion	6	1.194	0.331	1.57E-04	1
Curated gene sets: Croonquist NRAS signalling up	33	0.547	0.152	1.62E-04	1
GO BP: response to isolation stress	6	1.050	0.292	1.62E-04	1
GO BP: actin filament organization	370	0.154	0.044	2.26E-04	1
Curated gene sets: reactome aspartate and asparagine metabolism	11	0.995	0.286	2.48E-04	1
Curated gene sets: Boyault liver cancer subclass G3 DN	51	0.408	0.118	2.78E-04	1
GO MF: calcium dependent protein binding	59	0.367	0.109	3.87E-04	1
GO BP: alpha amino acid biosynthetic process	59	0.366	0.109	4.00E-04	1

Using FUMA, curated gene sets (C2.all) and gene ontology (GO) terms (C5. BP- Biological Processes, CC- Cellular Component and MF: Molecular Functions ontologies) are analysed using the MAGMA tool [84,211]. N genes- number of genes in the gene set, Beta- regression coefficient of the gene set, SE- standard error of the regression coefficient, P bon- P value after Bonferroni correction for multiple testing for all tested gene sets.

Figure 5.5 shows the MAGMA tissue expression analysis based on GTEXv8, which showed top genes associated with any serious irAE have a high expression at the left ventricle of the heart (P=0.04), followed by the transverse and sigmoid colon (P=0.05 and 0.09 respectively).

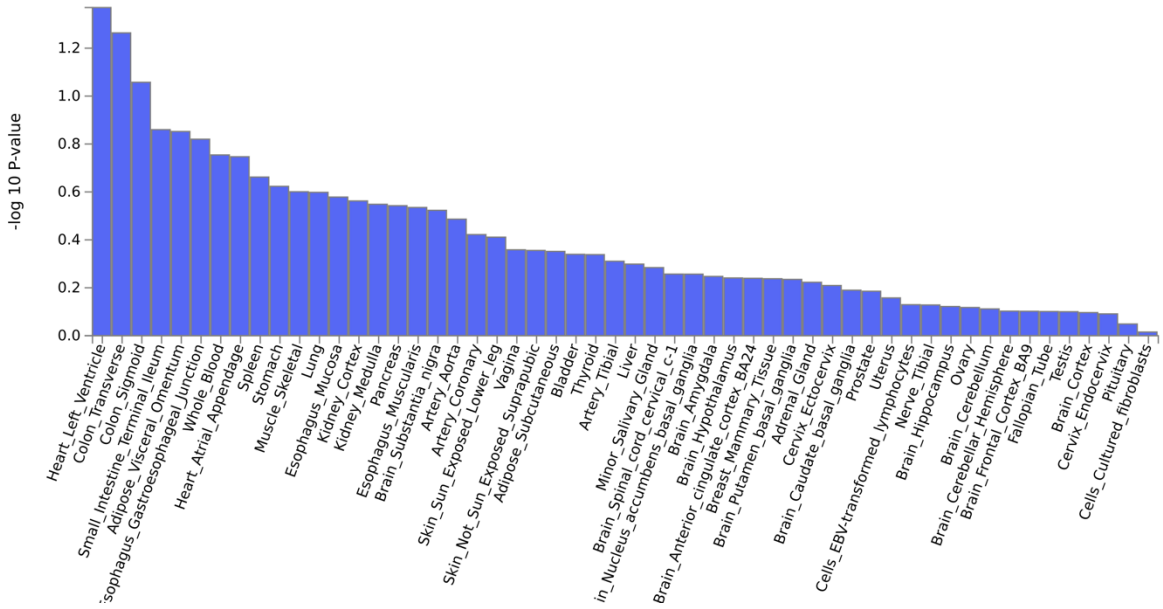


Figure 5.5 MAGMA tissue expression analysis of GWAS of serious irAE

MAGMA gene-property analysis is performed using FUMAv1.1.0[84,211]. Gene expression values are log2 transformed average RPKM per tissue type based on GTExv8 RNA-seq data[173]. Tissue expression analysis is performed separately for 53 specific tissue types for 17338 genes. Abbreviations: RPKM – reads per kilobase of exon per million reads mapped

Table 5.5 shows the eQTL data for the top SNPs associated with the presence of any serious irAE[84,173,175]. The most significant SNP-gene pair before adjusting for multiple testing include cis eQTL rs35429660G>C, which was associated with decreased in expression of *ZNF701* in the blood. This could potentially have a downstream impact on immune cells in the bloodstream that may influence the immune response to ICI. Another SNP that may have biological relevance in affecting immune response is rs7197466G>T which is associated with increased expression of *ERCC4* gene in T cells ($P_{corrected}=3.4x10^{-3}$), which are the main immune cells activated from ICI leading to infiltration and killing of cancer cells [27,237].

Table 5.5 Expression quantitative trait loci (eQTL) data from top SNPs associated with the presence of any serious irAE

SNP rsID	Risk allele	Gene	Aligned Direction	Tissue	P value	FDR	eQTL database
rs35429660	C	ZNF701	-	Blood	2.38E-51	0	eQTLGen
	C	ZNF701	+	Artery-Tibial	3.14E-10	1.60E-08	GTEx/v8
	C	ZNF701	+	Artery-Aorta	2.04E-07	1.42E-04	GTEx/v8
rs7197466	T	ERCC4	+	T-cell	6.74E-07	3.37E-03	eQTLcatalogue (BLUEPRINT)
rs4527728	T	CTA-212D2.2	-	Adipose-Subcutaneous	1.86E-05	5.01E-27	GTEx/v8
	T	PDE7B	-	Testis	6.47E-08	1.73E-04	GTEx/v8
rs2472135	A	RP11-21C17.1	-	Artery- Aorta	4.54E-05	2.03E-14	GTEx/v8
	A	RP11-600K15.1	-	Artery- Aorta	1.61E-05	6.09E-07	GTEx/v8
	A	GS1-44D20.1	-	Testis	3.80E-06	3.07E-05	GTEx/v8
	A	SULF1	-	Artery- Aorta	9.83E-06	5.49E-05	GTEx/v8

Unique pairs of SNP-tissue specific gene expression associations are shown for the top independent SNPs associated with any serious irAE. eQTL mapping was performed using FUMA based on the data sources listed[84]. Aligned direction is the direction of effect on gene expression, where "+" (risk allele increases gene expression) or "-" (risk allele decreases gene expression). FDR- False discovery rate of eQTLs after multiple testing. eQTL database- data source of eQTLs used: eQTLcatalogue (BLUEPRINT dataset includes immune cells based on 554 samples and 197 donors)[237], GTEx/v8[210]

Figure 5.6 shows the circos plot generated from chromatin interaction mapping for chromosome 19 that contains a risk variant (chr19:53088882) identified from the GWAS of irAE. Four genes, *ZNF83, ZNF701, ZNF808,* and *ZNF818P* (highlighted in red), are mapped by both chromatin interaction and eQTLs. The most significant intra-chromosomal interaction in this genomic locus between two regions, region 1: 19:53080001-53120000 (containing the genomic risk loci) and region 2: 19:53160001-53200000 (mapping associated genes) occurs within the mesenchymal stem cell (FDR of the interaction: $2.8x10^{-110}$, Hi-C data from GSE87112 database)[238].

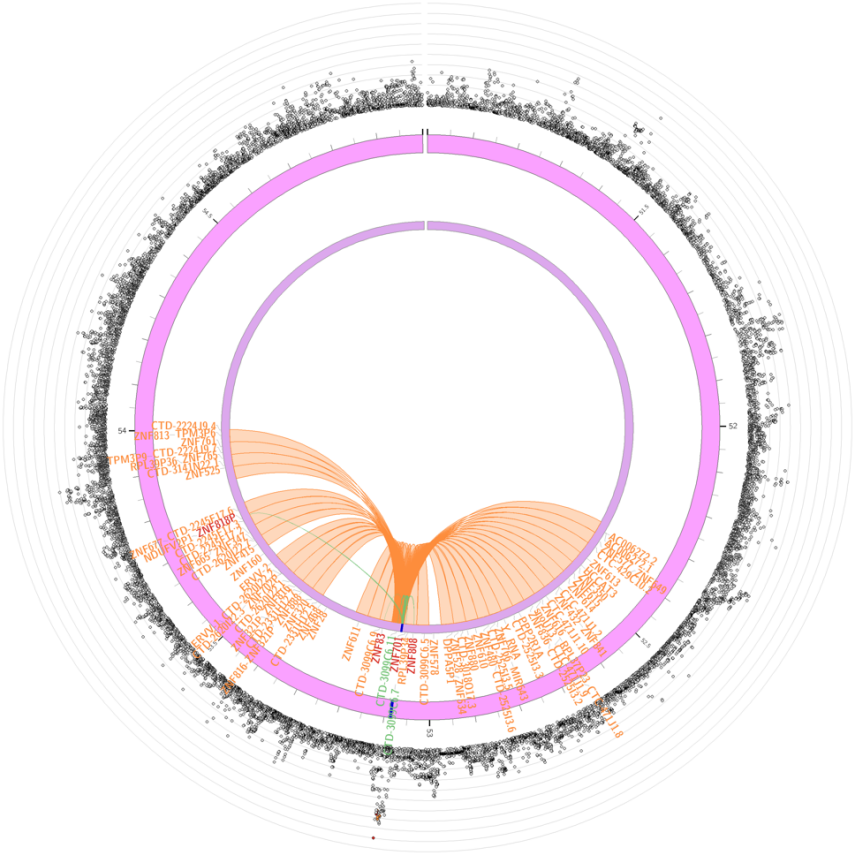


Figure 5.6 Chromatin interaction mapping for chromosome 19 based on GWAS of serious irAE
Build-in chromatin interaction mapping was performed using FUMA. The most outer layer of the circos plot is the Manhattan plot, where only SNPs with P<0.05 are displayed. Red dots are SNPS with r2>0.8 and grey dots are SNPS which are not in LD (r2 ≤0.2) with any independent significant SNPs in the locus. The second layer is the chromosome ring, where genomic risk loci are coloured in blue. Genes (in orange) are mapped by chromatin interaction, and genes (in green) are mapped by eQTLs only. Genes (in red) are mapped by both chromatin interaction and eQTL. The third layer is the chromosome ring which is the same as the second layer but without coordinates. Chromatin interaction links are coloured in orange, and eQTL links are coloured in green.

5.2.2.5 Association analysis of individual organ irAEs based on additive genetic models

Association testing using the frequentist approach to identify genetic variants linked with organ-specific ≥grade 3 irAE was performed based on the additive model and adjusted for co-variates age, sex, presence of concurrent cancer treatment and treatment with single or combination ICI therapy. Gastrointestinal (GI) irAE were the most common ≥grade 3 organ-specific irAE experienced by patients in the discovery cohort. These events include ICI-induced diarrhoea, colitis or gastritis events. Out of 186 patients with ≥grade 3 irAE, 75 (40%) were related to GI toxicities. 13% of the total study cohort (75/599) experienced a ≥grade 3 GI irAE.

Figure 5.7 shows the Manhattan plot of the results of the GWAS of GI irAE. None of the SNPs reached genome-wide significance in the analysis. **Table 5.6** details the top seven independent SNPs associated with the ≥grade 3 GI irAE that reached a P value threshold of $P<5x10^{-6}$. The top SNP identified from the GWAS is rs112264014(ins TTTG), which has an effect size of OR 10.92, 95% CI 4.32-27.61, $P=5.13x10^{-7}$ with a minor allele frequency (MAF) of 2% and mapped to the *MAST3* gene.

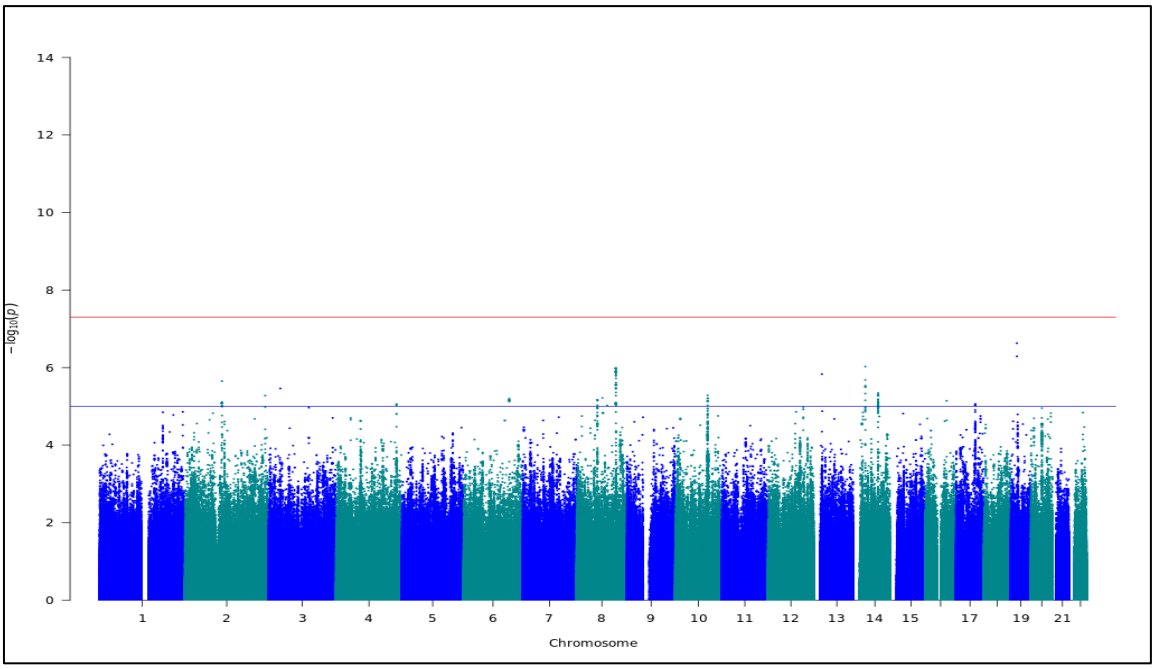


Figure 5.7 Manhattan plot from the genome-wide association analysis identifying variants predictive of ≥grade 3 GI irAE

The x-axis represents the SNPs from different chromosomes and the y-axis represents the significance level of association at -log10 of the P value for each SNP with ≥grade 3 GI events. Abbreviations: GI- gastrointestinal, irAE- immune-related adverse event

Table 5.6 Top independent SNPs identified from GWAS of ≥grade 3 GI irAE under the frequentist additive genetic model

Toxicity	rsID	Chr	Pos	Risk allele	Freq	INFO	OR	Lower CI	Upper CI	P value	Consequence	Mapped /nearest gene	RegulomeDB
≥Grade 3 GI events	rs112264014	19	18247619	TTTG	0.02	0.97	10.92	4.32	27.61	5.13E-07	intronic	MAST3	NA
	rs9552805	13	23607491	C	0.33	1	2.53	1.72	3.72	1.48E-06	intergenic	LINC00621(dist=116983),SGCG(dist=147569)	6
	rs145650938	14	34997739	T	0.06	0.77	6.43	3.03	13.62	2.08E-06	intronic	EAPP	7
	rs201085195	2	107745362	A	0.02	0.95	13.68	4.85	38.59	2.24E-06	intergenic	ST6GAL2(dist=241799),MIR548AU(dist=20758)	NA
	rs28481615	8	113972011	C	0.04	0.98	6.90	3.08	15.43	2.88E-06	intronic	CSMD3	5
	no rsID	3	34716436	A	0.91	0.59	28.39	3.98	202.35	3.44E-06	intergenic	PDCD6IP(dist=805237),LOC101928135(dist=200853)	NA
	rs35651129	14	72473744	CA	0.16	0.98	2.98	1.88	4.73	4.53E-06	intronic	RGS6	NA

Conditional and joint ('COJO') analysis was used to identify the top SNPs associated with ≥grade 3 GI irAE that reached a P-value threshold of $<5x10^{-6}$ after filtering for MAF>1% in the cases and control groups, and adjusting for covariates age, sex, single/doublet ICI therapy and presence of concurrent anti-cancer treatment. Grade 3-4 events were coded as cases and grade 0-2 events as controls. The risk allele refers to the allele associated with toxicity risk. Variant annotation was performed using the gene annotation software tool ANNOVAR. RegulomeDB scores (from 1a to 7) were obtained from RegulomeDBv1.1 using FUMA GWAS; the lowest score represents SNPs with the most biological evidence as a regulatory element [236][84].Abbreviations: Abbreviations: Chr- chromosome, CI- confidence interval (95%), Freq- frequency of risk allele, GI- gastrointestinal, INFO- info score, irAE- immune-related adverse event, MAF- minor allele frequency, OR- odds ratio, SNP- single nucleotide polymorphism, Pos- genomic position (build GRCh37)

Table 5.7 shows the top gene sets from the MAGMA gene-set analysis. The most significant associated gene ontology set was annotated to the regulation of alkaline phosphatase activity ($P=5.2x10^{-5}$, N=9 genes). The gene ontology set "positive regulation of myeloid leukocytes differentiation" was associated with ≥grade 3 GI irAEs ($P=8.8x10^{-4}$), but this was not significant after Bonferroni correction[172,211].

Table 5.7 Top 10 gene sets from MAGMA gene-set analysis for GWAS of ≥grade 3 GI toxicity

Gene Sets	N genes	Beta	SE	P value	P bon
GO BP: regulation of alkaline phosphatase activity	9	1.225	0.316	5.19E-05	0.8
GO BP: regulation of long-term synaptic potentiation	42	0.467	0.138	3.54E-04	1
GO MF: steroid hormone receptor binding	89	0.300	0.091	5.22E-04	1
GO BP: intracellular steroid hormone receptor signalling pathway	126	0.250	0.078	7.11E-04	1
GO MF: nuclear receptor binding	108	0.261	0.083	7.89E-04	1
GO BP: negative regulation of substrate adhesion dependent cell spreading	13	0.842	0.266	7.93E-04	1
GO BP: positive regulation of myeloid leukocyte differentiation	53	0.378	0.121	8.82E-04	1
GO BP: acrosomal vesicle exocytosis	5	1.029	0.332	9.77E-04	1
Curated gene sets: Kurokawa liver cancer early recurrence	6	1.136	0.370	1.07E-03	1
GO BP: small nucleolar RNA metabolic process	4	1.021	0.335	1.14E-03	1

Using FUMA, the curated gene sets (C2.all) and gene ontology (GO) terms (C5. BP- Biological Processes, CC- Cellular Component and MF: Molecular Functions ontologies) were analysed[84,211]. N genes- number of genes in the data in the gene set, Beta- regression coefficient of the gene set, SE- standard error of the regression coefficient, P bon- P value after Bonferroni correction for multiple testing for all gene sets tested.

Figure 5.8 shows that the top genes associated with GI toxicity were highly expressed in the main tissues of the GI tract, including the terminal ileum of the small intestine (P=0.03), transverse colon (P=0.04) and stomach (P=0.07).

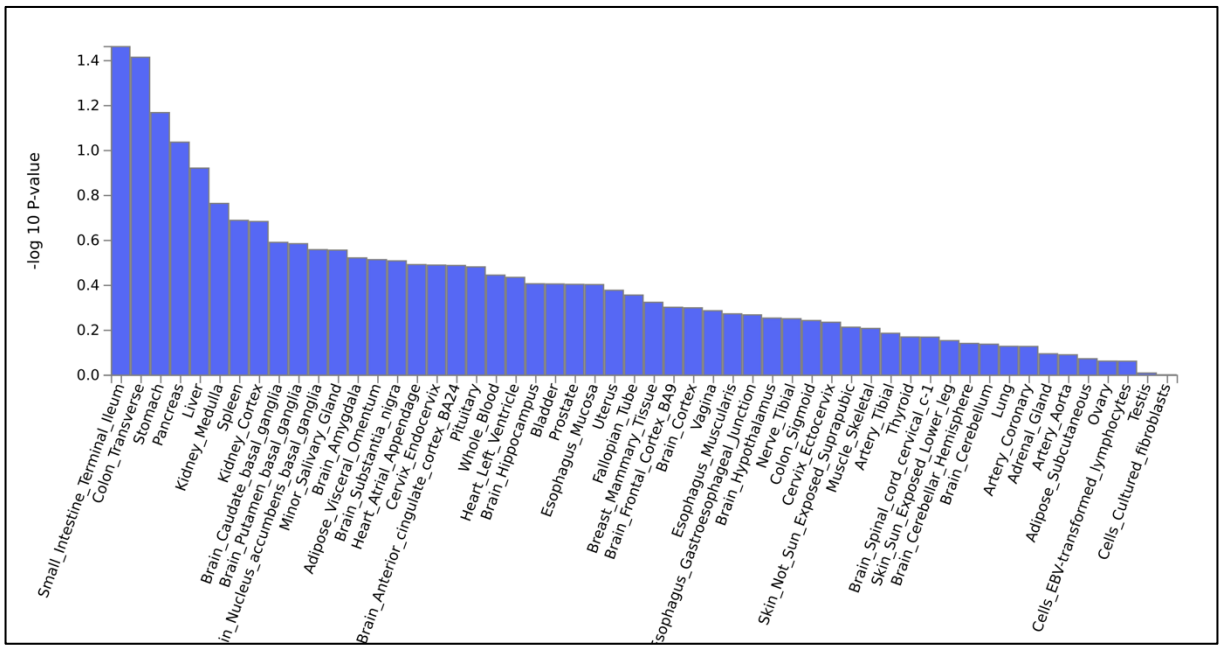


Figure 5.8 MAGMA tissue expression analysis from GWAS of ≥grade 3 GI toxicity
MAGMA gene-property analysis is performed via FUMAv1.1.0 using the average expression of genes per tissue type as a gene covariate [84,211]. Gene expression values are log2 transformed average RPKM per tissue type based on GTExv8 RNA-seq data. Tissue expression analysis is performed separately for 53 specific tissue types for 17,333 genes. Abbreviations: RPKM – reads per kilobase of exon per million reads mapped.

Table 5.8 shows the eQTL data for the top SNPs associated with the development of ≥grade 3 GI toxicity. One of the top cis eQTLs is rs145650938C>T, whereby the toxicity-associated allele is associated with lower expression of *SNX6* ($P=9.9x10^{-30}$) in the blood which may have biological relevance on systemic responses on different organ systems following ICI treatment.

Table 5.8 Expression quantitative trait loci (eQTL) data of the top SNPs associated with the presence of ≥grade 3 GI irAE

SNP	Risk allele	Gene	Aligned Direction	Tissue	P value	FDR	eQTL database
rs112264014	TTTG	UPF1	-	Sensory neurons	3.25E-06	1.62E-02	eQTLcatalogue (Schwartzentruber 2018)
rs145650938	T	SNX6	-	Blood	9.89E-30	0	eQTLGen
	T	KIAA0391	-	Blood	5.41E-08	2.21E-04	eQTLGen
	T	NFKBIA	-	Blood	8.55E-06	2.41E-02	eQTLGen
rs28481615	C	RP11-709P2.1	+	Pancreas	1.08E-06	1.08E-17	GTEx/v8
	C	CSMD3	+	Pancreas	4.77E-06	1.55E-16	GTEx/v8

Unique pairs of SNP-gene-tissue are shown for the top independent SNPs associated with ≥grade 3 GI irAE are shown. eQTL mapping was performed using FUMA GWAS based on the data sources listed[84]. Aligned direction is the direction of effect on gene expression, where "+" (risk allele increases gene expression) or "-" (risk allele decreases gene expression). FDR- False discovery rate of eQTLs after multiple testing. eQTL database- data source of eQTLs used: eQTLGen[175], eQTLcatalogue (Schwartzentruber dataset includes sensory neurons based on 98 samples and 98 donors)[239], GTEx/v8[210]

When considering the individual organ-specific irAEs alone, immune-related hepatitis had the highest rate of ≥grade 3 events, with 55/95 (58%) patients experiencing a ≥grade 3 event. **Figure 5.9** shows the results of the GWAS of hepatitis irAE in a Manhattan plot. There were no SNPs that reached genome-wide significance in the analysis. **Table 5.9** details the top 12 independent SNPs associated with the ≥grade 3 hepatitis irAE that reached a P value threshold of $P<5x10^{-6}$. rs61770027 (A>G) is the top SNP associated with ≥grade 3 irAE (OR 8.09 (95% CI 2.91-22.49), $P= 7.56x10^{-7}$). It is an intergenic SNP; nearest genes are *IFI44* and *ADGRL4.*

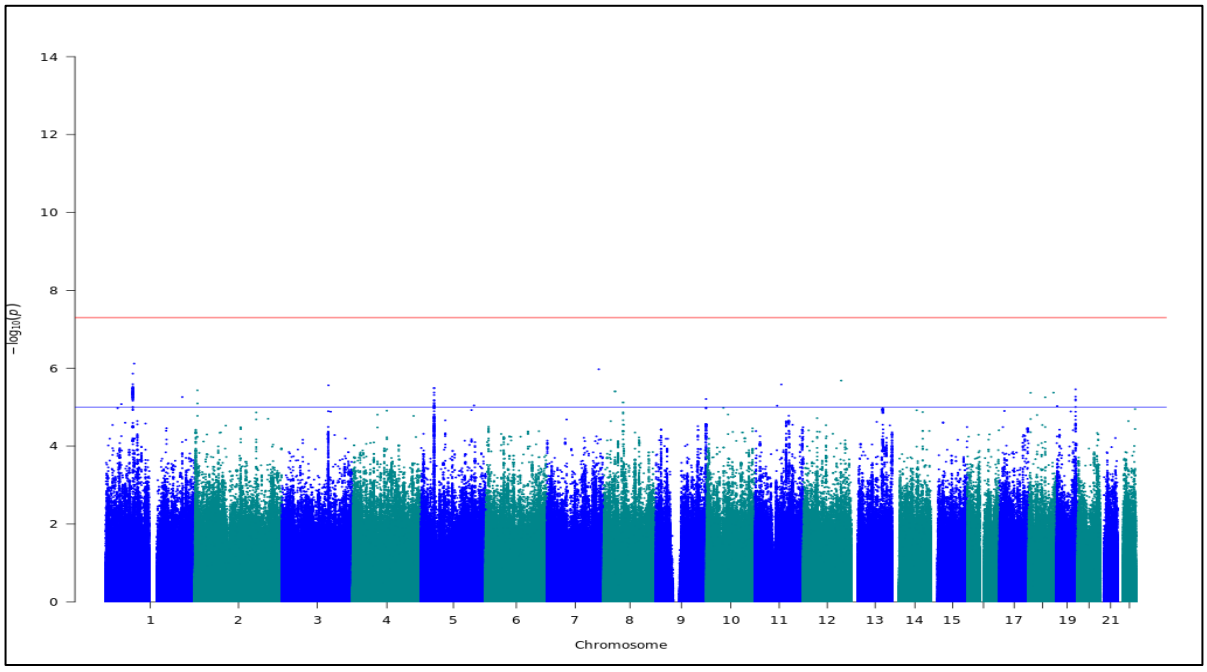


Figure 5.9 Manhattan plot from the genome-wide association analysis identifying variants predictive of ≥grade 3 immune-related hepatitis

The x-axis represents SNPs from different chromosomes and the y-axis represents the significance level of association at -log10 of the P value of each SNP with ≥grade 3 immune-related hepatitis is shown in the y-axis.

Table 5.9 Top independent SNPs identified from GWAS of ≥grade 3 hepatitis irAE under the frequentist additive genetic model

Toxicity	rsID	Chr	Pos	Risk allele	Freq	INFO	OR	Lower CI	Upper CI	P value	Consequence	Mapped/ nearest genes	Regulome DB
≥Grade 3 hepatitis	rs61770027	1	79175832	G	0.80	0.75	8.09	2.91	22.49	7.56E-07	intergenic	IFI44(dist=46069),ADGRL4(dist=179617)	6
	rs6974853	7	145121057	T	0.10	0.95	4.61	2.53	8.41	1.06E-06	intergenic	TPK1(dist=587911),CNTNAP2(dist=692396)	7
	rs1157858353	1	75433168	T	0.23	0.96	3.68	2.14	6.33	1.38E-06	intergenic	TYW3(dist=200808),LHX8(dist=160950)	NA
	rs4246254	12	106141104	T	0.02	1	13.89	4.69	41.16	2.08E-06	intergenic	CASC18(dist=3263),NUAK1(dist=316021)	6
	rs550587077	11	73539272	CT	0.86	0.82	18.29	2.93	114.28	2.61E-06	intronic	MRPL48	NA
	rs77941834	3	129956110	T	0.78	0.78	5.17	2.29	11.68	2.74E-06	ncRNA intronic	COL6A4P2	6
	rs4072644	5	36522484	T	0.78	1	4.63	2.14	9.98	3.23E-06	intergenic	RANBP3L(dist=220473),SLC1A3(dist=83973)	5
	rs6509652	19	53080873	C	0.52	0.93	2.95	1.81	4.82	3.47E-06	intronic	ZNF701	7
	rs7571339	2	6936488	C	0.42	0.85	2.94	1.83	4.74	3.69E-06	intergenic	LINC00487(dist=26046),NRIR(dist=32197)	6
	rs4457297	8	31488655	C	0.56	0.94	3.15	1.87	3.15	3.91E-06	intergenic	WRN(dist=455268),NRG1(dist=8613)	6
	rs11659272	18	69362897	T	0.07	0.65	9.47	3.70	24.24	4.23E-06	intergenic	LINC01541(dist=116705),LINC01899(dist=36955)	7
	rs71368789	18	5375886	T	0.22	0.98	3.33	1.98	5.62	4.27E-06	intergenic	ZBTB14(dist=78834),EPB41L3(dist=16389)	NA

Conditional and joint ('COJO') analysis was used to identify the top SNPs associated with ≥grade 3 immune-related hepatitis that reached a P-value threshold of $<5x10^{-6}$ after filtering for MAF>1% in the cases and control groups, and adjusting for covariates age, sex, single/doublet ICI therapy and presence of concurrent anti-cancer treatment. Grade 3-4 events were coded as cases and grade 0-2 events as controls. The risk allele refers to the allele associated with toxicity risk. Variant annotation was performed using the gene annotation software tool ANNOVAR. RegulomeDB scores (from 1a to 7) were obtained from RegulomeDBv1.1 using FUMA GWAS; the lowest score represents SNPs with the most biological evidence as a regulatory element [84,236]. Abbreviations: Chr- chromosome, CI- confidence interval (95%), Freq- frequency of risk allele, irAE- immune-related adverse event, MAF- minor allele frequency, ncRNA- non-coding RNA, OR- odds ratio, SNP- single nucleotide polymorphism, Pos- genomic position (build GRCh37)

Table 5.10 shows that the top significant gene ontology sets include nicotinamide adenine dinucleotide (NAD) biosynthetic process from tryptophan (OR=4.21, P=2.4x10^{-5}, N=7 genes), tryptophan metabolic process (OR= 3.34, P=6.2x10^{-5}, N=8 genes) and negative regulation of the retinoic acid-inducible gene I (RIG-I) stimulating pathway (OR=2.63, P=1.5x10^{-4}, N=8 genes). Another significant gene set includes the genes upregulated in five endothelial cell types by TNFα (OR=1.37, P=5.8x10^{-4}, N=82 genes), which is a stimulatory factor and plays a role in cancer antigen recognition as part of the cancer immunity cycle[27]. None of the gene sets were associated with hepatitis with statistical significance after Bonferroni correction.

Table 5.10 Top 10 gene sets from MAGMA gene-set analysis for GWAS of ≥grade 3 hepatitis

Gene Sets	N genes	Beta	SE	P value	P bon
GO BP: de novo NAD biosynthetic process from tryptophan	7	1.438	0.354	2.40E-05	0.372
GO BP: tryptophan metabolic process	8	1.207	0.314	6.20E-05	0.960
GO BP: negative regulation of RIG-I signalling pathway	8	0.968	0.268	1.51E-04	1
GO BP: prostanoid metabolic process	14	0.698	0.200	2.39E-04	1
GO BP: antifungal innate immune response	4	1.059	0.313	3.61E-04	1
Curated gene sets: Browne HCMV infection 48hr	444	0.138	0.041	3.84E-04	1
Curated gene sets: Hellebrekers silenced during tumour angiogenesis	75	0.329	0.099	4.51E-04	1
GO BP: negative regulation of viral release from the host cell	14	0.718	0.218	4.81E-04	1
Curated gene sets: Staege Ewing family tumour	26	0.555	0.170	5.66E-04	1
Curated gene sets: Sana TNF signalling up	82	0.317	0.098	5.81E-04	1

Using FUMA, the curated gene sets (C2.all) and gene ontology (GO) terms (C5. BP- Biological Processes, CC- Cellular Component and MF: Molecular Functions ontologies) are analysed[84,211]. N genes- number of genes in the data in the gene set, Beta- regression coefficient of the gene set, SE- standard error of the regression coefficient, P bon- P value after Bonferroni correction for multiple testing for all tested gene sets. Abbreviations: NAD- nicotinamide adenine dinucleotide, RIG-I- retinoic acid-inducible, HCMV- human cytomegalovirus

The top genes identified from the GWAS of immune-related hepatitis were most highly expressed in tissues including the vagina (P=0.003), oesophagus (P=0.005) and transverse colon (P=0.012) but also showed strong expression in liver tissue (P=0.119) (**Figure 5.10**).

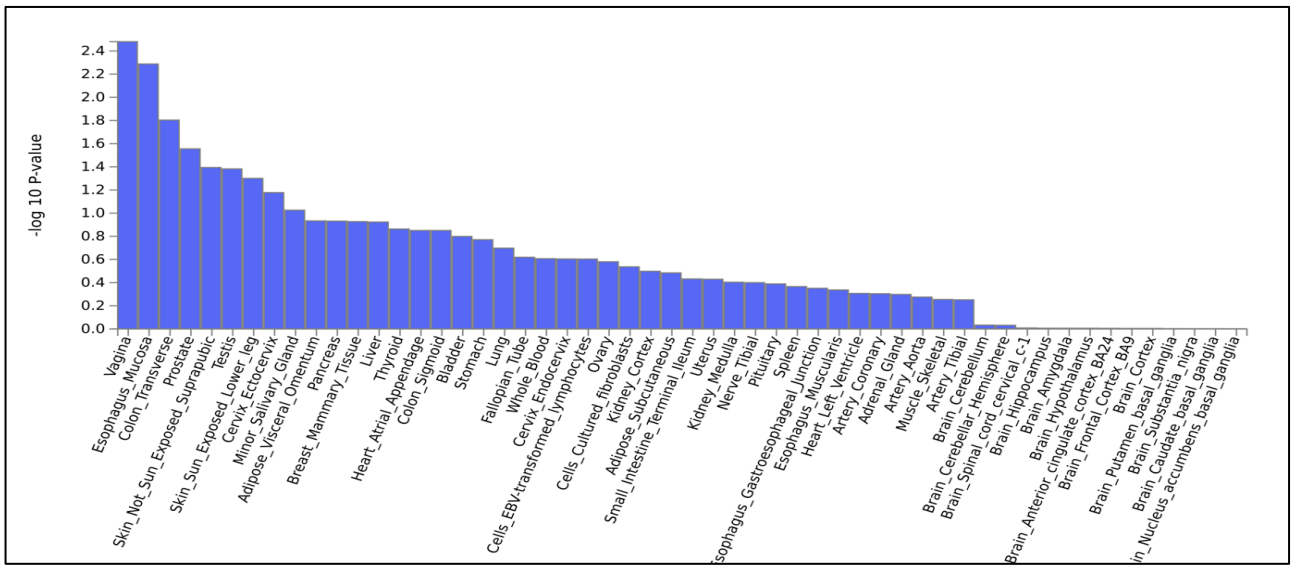


Figure 5.10 MAGMA tissue expression analysis of GWAS of ≥grade 3 immune-related hepatitis
MAGMA gene-property analysis is performed via FUMAv1.1.0 using the average expression of genes per tissue type as a gene covariate [84,211]. Gene expression values are log2 transformed average RPKM per tissue type based on GTExv8 RNA-seq data[173]. Tissue expression analysis is performed separately for 53 specific tissue types for 17331 genes. Abbreviations: RPKM – reads per kilobase of exon per million reads mapped

Table 5.11 shows the eQTL data for the top SNPs associated with the development of ≥grade immune-related hepatitis. Two of the top SNP-gene pairs including rs77941834 and rs6509652 control the expression of *RP11-77P16.4* and *ZNF701* respectively in T cells, which may influence the immune response of patients receiving ICI treatment and subsequently the development of immune-related toxicity.

Table 5.11 Expression quantitative trait loci (eQTL) data of top SNPs associated with the presence of ≥grade 3 immune-related hepatitis

SNP	Risk allele	Gene	Aligned Direction	Tissue	P value	FDR	eQTL database
rs77941834	T	RP11-77P16.4	-	LCL	5.04E-06	2.52E-02	eQTLcatalogue (GENCORD)
	T	RP11-77P16.4	-	T-cell	7.73E-08	3.87E-04	eQTLcatalogue
	T	RP11-77P16.4	-	LCL	2.95E-13	1.48E-09	eQTLcatalogue (GEUVADIS)
	T	RP11-77P16.4	+	Adrenal gland	2.60E-07	1.96E-34	GTEx/v8
	T	RP11-77P16.4	+	EBV-transformed lymphocytes	1.51E-05	1.68E-37	GTEx/v8
	T	RP11-77P16.4	+	Artery- Aorta	1.31E-13	9.71E-59	GTEx/v8
	T	COL6A4P2	+	Adipose- Visceral omentum	3.67E-12	4.70E-36	GTEx/v8
	T	COL6A4P2	+	Artery- Tibial	1.33E-13	1.53E-72	GTEx/v8
	T	ALG1L2	+	Brain- Nucleus accumbens basal ganglia	1.43E-05	3.36E-09	GTEx/v8
	T	COL6A4P2	+	Brain- Putamen basal ganglia	2.21E-09	8.72E-27	GTEx/v8
	T	COL6A4P2	+	Breast- Mammary tissue	3.08E-05	2.43E-30	GTEx/v8
	T	RP11-77P16.4	+	Breast-Mammary tissue	9.17E-05	2.80E-24	GTEx/v8
	T	COL6A4P2	+	Colon- Sigmoid	9.05E-10	9.31E-40	GTEx/v8
	T	RP11-77P16.4	+	Colon- Sigmoid	2.84E-10	8.29E-33	GTEx/v8
	T	COL6A4P2	+	Oesophagus- GOJ	1.30E-10	3.22E-41	GTEx/v8
	T	RP11-77P16.4	+	Oesophagus- GOJ	1.31E-08	3.25E-26	GTEx/v8
	T	COL6A4P2	+	Heart- Atrial appendage	4.31E-06	1.23E-26	GTEx/v8
	T	RP11-77P16.4	+	Heart- Left ventricle	2.89E-07	6.27E-44	GTEx/v8
	T	COL6A4P2	+	Lung	1.01E-10	4.91E-51	GTEx/v8
	T	RP11-77P16.4	+	Lung	5.03E-09	2.17E-40	GTEx/v8
	T	RP11-77P16.4	+	Muscle- Skeletal	1.17E-04	6.27E-25	GTEx/v8
	T	COL6A4P2	+	Nerve- Tibial	4.82E-07	1.02E-48	GTEx/v8
	T	COL6A4P2	+	Pancreas	7.69E-06	4.89E-30	GTEx/v8
	T	RP11-77P16.4	+	Pancreas	5.32E-05	7.38E-20	GTEx/v8
	T	COL6A4P2	+	Pituitary	5.08E-08	1.99E-30	GTEx/v8
	T	RP11-77P16.4	+	Pituitary	3.98E-07	8.95E-38	GTEx/v8
	T	COL6A4P2	+	Prostate	1.45E-05	1.81E-22	GTEx/v8
	T	RP11-77P16.4	+	Prostate	1.14E-05	2.07E-17	GTEx/v8
	T	RP11-77P16.4	+	Fibroblasts	3.08E-12	3.48E-71	GTEx/v8
	T	RP11-77P16.4	+	Stomach	5.21E-05	5.68E-15	GTEx/v8
	T	RP11-77P16.4	+	Testis	1.54E-06	7.12E-28	GTEx/v8
	T	COL6A4P2	+	Thyroid	3.21E-16	3.29E-84	GTEx/v8
	T	RP11-77P16.4	+	Thyroid	1.72E-13	3.28E-66	GTEx/v8
rs4072644	T	SLC1A3	-	Thyroid	4.16E-06	2.25E-19	GTEx/v8
rs6509652	C	ZNF701	-	T-cell- CD4	3.48E-06	1.74E-02	eQTLcatalogue (CEDAR)
	C	ZNF83	+	Blood	2.05E-10	0	eQTLGen
	C	ZNF702P	+	Blood	2.32E-06	6.78E-03	eQTLGen
	C	ZNF701	+	Oesophagus mucosa	9.11E-07	1.54E-07	GTEx/v8

	C	ZNF701	+	Lung	4.83E-06	4.53E-07	GTEx/v8
	C	ZNF701	+	Thyroid	1.42E-05	2.76E-04	GTEx/v8
rs7571339	C	LINC00487	+	Brain	2.62E-07	1.31E-03	eQTLcatalogue (BrainSeq)
	C	LINC00487	-	Blood	9.09E-21	0	eQTLGen
	C	LINC00487	-	Adipose- Subcutaneous	5.40E-13	3.49E-20	GTEx/v8
	C	LINC00487	-	Whole Blood	5.12E-05	3.24E-12	GTEx/v8
	C	LINC00487	-	Brain- Caudate basal ganglia	9.07E-08	6.63E-07	GTEx/v8
	C	LINC00487	-	Breast- Mammary tissue	2.82E-10	3.71E-13	GTEx/v8
	C	LINC00487	-	Oesophagus mucosa	2.80E-06	1.19E-11	GTEx/v8
	C	LINC00487	-	Heart-Left ventricle	1.03E-06	5.97E-11	GTEx/v8
	C	LINC00487	-	Lung	2.94E-07	2.28E-25	GTEx/v8
	C	LINC00487	-	Muscle- Skeletal	4.00E-12	3.42E-27	GTEx/v8
	C	LINC00487	-	Nerve- Tibial	1.19E-07	6.91E-19	GTEx/v8
	C	LINC00487	-	Testis	1.25E-09	2.26E-12	GTEx/v8
	C	LINC00487	-	Thyroid	4.71E-08	2.11E-15	GTEx/v8

Unique pairs of SNP-tissue specific gene expression are shown for the top independent SNPs associated with ≥grade 3 immune-related hepatitis toxicity. eQTL mapping was performed using FUMA GWAS based on the data sources listed[84]. Aligned direction is the direction of effect on gene expression, where "+" (risk allele increases gene expression) or "-" (risk allele decreases gene expression). FDR- False discovery rate of eQTLs after multiple testing. eQTL database- data source of eQTLs used: eQTLGen[175], eQTLcatalogue (GENCORD based on 560 samples and 195 donors [240], GEUVADIS based on 445 samples and donors [241], CEDAR based on 2338 samples and 322 donors[242] and BrainSeq based on 484 samples and donors[243]), GTEx/v8[210]. For SNP-gene pairs with multiple tissue subtypes, the most significant paired tissue subtype is included. Abbreviations: EBV- Epstein-Barr virus, GOJ- gastro-oesophageal junction, LCL- lymphoblastoid cell line

Table 5.12 shows the top SNPs identified for other reported irAE, including endocrine, pneumonitis, skin and musculoskeletal toxicities that reached suggestive levels of significance ($P<5x10^{-6}$). The endocrine toxicity variable included ICI-induced thyroiditis or hyperglycaemia cases. Due to the insufficient number of ≥grade 3 cases causing lack of power, GWAS to identify variants predictive of ≥grade 2 skin, musculoskeletal toxicities and pneumonitis were performed. All three toxicity phenotypes have less than 20 ≥grade 3 cases, thus would only have been adequately powered at >80% to detect associated common variants (MAF>30%) with large effect sizes (OR >2.5) and rare variants (MAF<10%) at OR >2.9.

12 SNPs were associated with ≥grade 3 endocrine irAE, 7 SNPs associated with ≥grade 2 pneumonitis, 10 SNPs associated with ≥grade 2 skin irAE, and 14 were SNPs associated with ≥grade 2 musculoskeletal irAEs. Most variants are common with MAFs of >30% with low to moderate effect sizes, although several SNPs associated with endocrine and MSK irAEs are rare with MAFs of <10% with large effect sizes at OR>10. **Table 5.13** shows the significant eQTL data for the top SNPs identified using the GTEXv8 portal. The SNP-gene pairs are expressed in multiple tissue types. Several significant eQTL associated with cutaneous irAE are expressed in skin tissue.

Table 5.12 Top independent SNPs identified from GWAS of ICI-induced endocrine, pneumonitis, skin and musculoskeletal toxicities under the frequentist additive genetic model

Toxicity	rsID	Chr	Pos	Risk allele	Freq	INFO	OR	Lower CI	Upper CI	P value	Consequence	Mapped gene/nearest genes
Endo (012v34)	rs2237988	11	17444134	T	0.246	0.99	4.21	2.40	7.37	2.34E-07	intronic	ABCC8
	rs4836205	5	125018733	A	0.685	0.98	7.14	2.57	19.81	6.51E-07	intergenic	LINC02240(dist=80813),LINC02039(dist=496537)
	rs9285450	6	123627922	T	0.189	0.99	3.97	2.29	6.86	8.40E-07	intronic	TRDN
	rs35938141	6	68050943	A	0.060	0.98	7.57	3.42	16.74	1.87E-06	intergenic	SLC25A51P1(dist=1551567),LOC102723883(dist=714300)
	rs6718493	2	5229400	A	0.697	0.99	5.73	2.35	13.96	2.95E-06	intergenic	LINC01249(dist=525588),LINC01248(dist=544873)
	rs28759175	1	203495475	T	0.042	1	8.76	3.77	20.37	3.04E-06	intergenic	OPTC(dist=17398),ATP2B4(dist=100440)
	rs71484114	11	40717591	G	0.025	0.66	24.52	7.17	83.82	3.27E-06	intronic	LRRC4C
	rs12657662	5	163065926	T	0.656	0.98	5.01	2.21	11.36	3.62E-06	intergenic	MAT2B(dist=119567),LINC02143(dist=809502)
	rs8121883	20	50197080	T	0.726	1	6.46	2.31	18.09	3.98E-06	intergenic	NFATC2(dist=17710),ATP9A(dist=16234)
	rs7960857	12	118039445	C	0.187	0.95	4.15	2.26	7.64	4.14E-06	intronic	KSR2
	rs6425268	1	173961837	G	0.080	0.61	9.17	3.71	22.65	4.72E-06	intronic	RC3H1
	rs4654940	1	21787746	A	0.608	0.85	5.10	2.31	11.27	4.82E-06	intronic	NBPF3
Resp (01v234)	rs4742394	9	789299	T	0.733	0.97	23.01	2.81	188.73	8.21E-07	intergenic	FAM107B
	rs4526214	6	141777192	A	0.500	0.91	5.85	2.55	13.41	1.31E-06	intergenic	CASC2(dist=72829),FAM204A(dist=26078)

	rs2609822	10	14755798	G	0.408	0.98	4.78	2.33	9.80	3.34E-06	intronic	LINC00929(dist=47985),GABRB3(dist=362525)
	rs113647779	10	120042494	T	0.571	0.65	9.39	3.13	28.19	3.49E-06	intergenic	PAQR3
	rs34186526	5	29462333	T	0.096	0.97	7.12	3.11	16.33	3.60E-06	intergenic	LINC02064(dist=66250),LOC105374704(dist=418334)
	rs532856	15	26426169	T	0.558	0.78	6.59	2.58	16.82	4.01E-06	intergenic	MIR4465(dist=772172),NMBR(dist=618429)
	rs968706723	4	79845869	GT	0.140	0.91	4.97	2.56	9.62	4.38E-06	intronic	KANK1(dist=43193),DMRT1(dist=52391)
Skin (01v234)	rs368300917	12	116199923	A	0.81	1	3.55	1.90	6.64	4.19E-06	intergenic	TBX3(dist=1077954),MED13L(dist=196458)
	rs2328960	13	76176003	G	0.85	0.99	4.34	2.06	9.17	3.91E-06	intronic	UCHL3
	rs796605942	14	88537389	CA	0.92	0.66	23.15	3.42	156.43	3.68E-06	ncRNA_intronic	LINC01146
	rs12904674	15	31076481	C	0.85	1	5.29	2.38	11.75	6.40E-07	intergenic	LOC100288637(dist=11272),HERC2P10(dist=33758)
	rs199582124	18	13747231	TG	0.13	1	3.55	2.24	5.62	1.10E-07	intronic	RNMT
	rs483218	1	171608829	G	0.04	0.95	5.66	2.80	11.43	4.90E-06	intronic	MYOC
	rs139715	22	25271044	C	0.90	0.77	12.78	2.70	60.42	4.73E-06	intronic	SGSM1
	rs12483981	22	51142381	A	0.18	0.82	2.96	1.89	4.61	2.32E-06	intronic	SHANK3
	rs4334749	4	14984760	T	0.63	0.99	2.46	1.63	3.70	4.08E-06	ncRNA_intronic	CPEB2-DT
	rs67402512	5	10123850	C	0.65	0.93	2.87	1.84	4.49	4.80E-07	intergenic	LINC02112(dist=219914),FAM173B(dist=101770)
MSK (01v234)	rs2038457	20	42239145	A	0.17	0.95	4.45	2.65	7.47	1.43E-08	intronic	IFT52
	rs820336	3	123415781	T	0.77	1	7.34	2.63	20.52	6.48E-07	intronic	MYLK
	rs1542396	17	77350232	C	0.70	0.94	5.07	2.32	11.06	6.99E-07	intronic	RBFOX3

	rs1420877	4	161707186	T	0.53	0.99	3.53	2.94	6.11	7.94E-07	intergenic	LINC02233(dist=1008250),FSTL5(dist=597858)
	rs77592590	11	72156569	T	0.03	0.80	13.88	5.28	36.51	1.06E-06	intergenic	CLPB(dist=10841),LINC01537(dist=125131)
	rs1022831	9	658360	A	0.82	1	11.98	2.70	53.06	1.13E-06	intronic	KANK1
	rs6498069	16	10630835	A	0.22	0.99	3.48	2.08	5.81	1.53E-06	intronic	EMP2
	rs180877139	2	134888265	G	0.02	0.79	15.98	5.64	45.24	1.98E-06	intergenic	MIR3679(dist=3502),MGAT5(dist=123565)
	rs11995251	8	96531284	C	0.83	0.97	11.44	2.57	50.94	2.50E-06	ncRNA_intronic	C8orf37-AS1
	rs55724343	7	113055008	T	0.03	0.87	9.70	4.02	23.40	3.53E-06	intergenic	SMIM30(dist=296371),PPP1R3A(dist=461874)
	rs139908045	2	122830740	T	0.02	0.90	11.33	4.39	29.22	4.32E-06	intergenic	TSN(dist=305312),LINC01826(dist=992992)
	rs13187338	5	164057994	G	0.80	0.97	7.29	2.32	0.84	4.69E-06	intergenic	LOC102546299(dist=88005),NONE(dist=NONE)
	no rsID	7	53333854	GTA	0.79	0.79	8.25	2.54	26.80	4.92E-06	intergenic	POM121L12(dist=229236),LINC01446(dist=389346)
	rs1249154	10	79979872	G	0.09	0.82	4.34	2.39	7.87	4.99E-06	intergenic	RPS24(dist=163301),LINC00856(dist=28510)

Conditional and joint ('COJO') analysis was used to identify the top SNPs associated with ICI-induced toxicities that reached a P-value threshold of $<5x10^{-6}$ after filtering for MAF>1% in the cases and control groups, and adjusting for covariates age, sex, single/doublet ICI therapy and presence of concurrent anti-cancer treatment. For the pneumonitis, skin and MSK toxicity variables, cases were defined as events of ≥grade 2 severity. The risk allele refers to the allele associated with toxicity risk. Abbreviations: Chr- chromosome, CI- confidence interval (95%), Endo- endocrine, Freq- frequency of risk allele, irAE- immune-related adverse event, MAF- minor allele frequency, MSK- musculoskeletal, ncRNA- non-coding RNA, OR- odds ratio, SNP- single nucleotide polymorphism, Pos- genomic position (build GRCh37)

Table 5.13 Expression quantitative trait loci (eQTL) data of top SNPs associated with the presence of ≥grade 3 endocrine and ≥grade 2 pneumonitis, skin and MSK irAE

Toxicity	SNP	Ref allele	Alt allele	Gene	P value	NES	Tissue
Endocrine (012v34)	rs28759175	C	T	ATP2B4	1.60E-04	0.18	Cells - Cultured fibroblasts
	rs6718493	C	A	SOX11	1.30E-06	-0.32	Heart
	rs2237988	T	C	ABCC8	4.60E-05	-0.26	Thyroid
	rs8121883	T	C	ATP9A	4.00E-06	-0.16	Brain
Respiratory (01v234)	rs4526214	G	A	RP11-63E9.1	1.80E-29	0.35	Muscle - Skeletal
	rs2609822	G	A	FAM107B	9.00E-14	0.27	Skin
Skin (01v234)	rs483218	C	G	DNM3	9.70E-07	0.49	Skin
				METTL13	6.40E-06	0.31	Skin
	rs4334749	C	T	LINC00504	4.10E-25	-0.4	Skin
	rs2328960	T	G	CPEB2	1.40E-17	0.28	Oesophagus
				RP11-665G4.1	1.40E-04	-0.12	Artery
				UCHL3	1.80E-22	0.32	Artery
				COMMD6	7.40E-08	0.35	Testis
				LMO7-AS1	1.50E-09	0.38	Adipose
				TBC1D4	2.10E-06	0.42	Brain
	rs12904674	C	T	ARHGAP11B	9.30E-05	-0.24	Muscle
				CHRFAM7A	1.90E-16	0.62	Thyroid
				CTD-3092A11.1	2.90E-09	-0.42	Skin
				CTD-3092A11.2	2.70E-05	0.25	Thyroid
				CTD-3092A11.3	5.10E-13	-0.55	Adipose
				FAN1	2.50E-20	-0.54	Brain
				GOLGA8H	3.80E-09	0.55	Heart
				GOLGA8Q	2.60E-10	-0.42	Thyroid
				GOLGA8R	7.70E-13	0.35	Whole Blood
				GOLGA8T	3.50E-08	0.36	Thyroid
				GOLGA8UP	2.10E-08	-0.63	Brain
				HERC2P10	2.10E-35	-1	Brain
				MTMR10	5.10E-05	0.14	Nerve

				RP11-540B6.2	7.00E-31	-1.1	Brain
				RP11-540B6.3	5.70E-16	-0.79	Brain
				RP11-540B6.6	2.90E-18	-0.6	Brain
				RP11-932O9.9	2.00E-06	-0.56	Prostate
				ULK4P2	3.40E-07	0.47	Heart
				ULK4P3	6.40E-20	0.69	Thyroid
MSK (01v234)	rs2038457	A	G	RP1-138B7.7	3.00E-41	-0.72	Testis
				IFT52	5.30E-08	0.21	Oesophagus
				MYBL2	2.30E-18	-0.44	Spleen
				TOX2	8.90E-05	-0.28	Nerve
	rs820336	C	T	MYLK	9.00E-05	0.1	Nerve
	rs1542396	T	C	CTD-2047H16.3	3.80E-05	0.29	Artery
	rs1022831	A	C	KANK1	1.90E-07	-0.17	Heart
	rs6498069	A	G	EMP2	5.80E-14	0.51	Pancreas
				LINC01290	5.20E-06	-0.29	Nerve
				RP11-609N14.4	9.20E-13	-0.24	Skin
	rs13187338	G	A	CTC-340A15.2	3.80E-06	0.31	Testis

Unique pairs of SNP-tissue specific gene expression are shown for the top independent SNPs associated with the indicated irAE. Gene expression was obtained from the GTEX v8 portal[210]. Only the most significant tissue-specific gene expression association is included in this table. Normalized effect size (NES) represents the slope of the linear regression, computed as the effect of the alternative (ALT) allele relative to the reference (REF) allele. Abbreviations: MSK- musculoskeletal

The cumulative OST score is an alternative toxicity grading system designed by the PAIR (The Pathobiology of Adverse Immune Reactions) study team from Kings College London. The PAIR study is a prospective study recruiting patients treated with ICI agents. The cumulative OST ("T") score is determined by considering the number of organs affected by toxicity ("O") and the number of CTCAE grade ≥3 toxicities ("S") demonstrated as below:

$$O + (S*3) = T$$

I performed a GWAS of a categorical variable generated from the patients' T scores. Two SNPs reached GWAS significance and are shown in **Table 5.14**.

Table 5.14 Top SNPs associated with cumulative OST scores under the additive model that reached GWAS significance (P x $5x10^{-8}$)

rsID	Chr	Pos	Risk allele	Freq	INFO	OR (95%CI)			P value (lrt)	Consequence	Mapped gene/nearest genes
						Low tox (n=193)	Moderate tox (n=149)	High tox (n=43)			
rs35429660	19	53088882	C	0.65	0.93	1.12 (0.83-1.52)	1.82(1.25-2.64)	7.52 (3.43-16.51)	1.30E-08	UTR3	ZNF701
rs6821356	4	31219796	C	0.65	1	2.45 (1.77-3.38)	2.36 (1.63-3.41)	2.13 (1.24-3.67)	4.73E-08	intergenic	LINC02497(dist=6500),LINC02501(dist=288370)

Association testing for categorical traits was performed to identify top SNPs associated with cumulative T scores using the multinomial logistic regression likelihood. A cumulative OST score was calculated for each patient sample based on the number of organs affected and the number of grade 3 events experienced with ICI therapy. Patients were grouped according to their scores (1-3: low toxicity,4-7: moderate toxicity, 8-14: high toxicity group). The risk allele refers to the allele associated with toxicity risk. P value was calculated based on the likelihood ratio test (lrt).Abbreviations: Chr- chromosome, CI- confidence interval (95%), Freq- frequency of risk allele OR- odds ratio, SNP- single nucleotide polymorphism, Pos- genomic position (build GRCh37), UTR- untranslated region.

Table 5.15 shows the P values of the top SNPs identified in the T score analysis when analysed in relation to the any serious irAE toxicity variable and the P values of the top SNPs associated with any serious irAEs when analysed in relation to the T score variable. rs35429660 was identified as a top SNP in both GWAS analyses.

Table 5.15 Performance of the top SNPs associated with serious irAE and OST score

Serious toxicity top hits	**P value in GWAS**	**P value in OST score analysis**
rs71399956	8.91E-08	1.81E-05
rs35429660	5.13E-07	1.30E-08
OST score top hits	**P value in GWAS**	**P value in serious toxicity analysis**
rs35429660	1.30E-08	5.13E-07
rs6821356	4.73E-08	2.40E-02

5.2.2.6 Association analysis based on Cox proportional hazard models

I next conducted a GWAS taking into account the amount of time taken to the development of an irAE using the Cox proportional hazards model. Unlike the logistic regression model used in SNPTEST for the previous GWAS which tests whether the toxicity risk variant affects the odds of toxicity, the Cox model tests whether the variant affects the onset time of the event. The Cox model may be more relevant in studies following up patients for the development of irAE for a period of time[244]. It takes into account the time to onset of the irAE, which can vary between different types of irAE and ICI treatment regimes and gives different weighting to early and late onset events[244]. This GWAS was based on 542/599 patients with time-to-toxicity data available for this analysis. **Table 5.16** shows the top SNPs identified in the GWAS analysis using the Cox proportional hazard ratio model. Seven SNPs were associated with the development of any serious irAE at the GWAS significance level ($P<5x10^{-8}$). The majority of the SNPs were rare with a MAF of 1%. The top SNP rs181554712 was associated with serious irAE

with a HR 7.69, 95% CI 3.83-15.42, P=9.43x10^{-9}) and maps to an intergenic region nearest to the *NEGR1* gene and long intergenic non-protein coding RNA LINC01360. *NEGR1* is found to be involved in cell adhesion processes and may function as a trans-neural growth promoting factor in regenerative axons but has no known associations with irAE[245].

Table 5.16 Top independent SNPs identified from GWAS of serious irAE under the Cox proportional hazard model

rsID	Chr	Pos	Effect allele	Freq	INFO	HR	Lower CI	Upper CI	P value	Consequence	Mapped gene/nearest genes
rs181554712	1	72931385	G	0.01	0.74	7.69	3.83	15.42	9.43E-09	intergenic	NEGR1(dist=183108), LINC01360(dist=840468)
rs185491919	5	41533622	T	0.01	0.88	10.02	4.47	22.47	2.24E-08	intergenic	PLCXD3(dist=22892),OXCT1 (dist=196545)
rs143143335	12	42771817	G	0.01	0.54	19.32	6.83	54.66	2.42E-08	intronic	PPHLN1
rs541040587	6	43535630	A	0.01	0.86	7.47	3.68	15.14	2.46E-08	intronic	XPO5
rs116463228	2	49584354	T	0.01	0.95	6.11	3.22	11.60	2.96E-08	intergenic	FSHR(dist=202688),NRXN1(dist=561289)
rs179897	7	106036872	C	0.99	0.74	0.12	0.06	0.25	3.60E-08	intergenic	NAMPT(dist=111234),CCDC 71L(dist=257936)
rs183868187	5	41940458	T	0.01	0.77	11.23	4.73	26.65	4.12E-08	intronic	FBX04

Conditional and joint ('COJO') analysis was used to identify the top SNPs associated with serious irAE that reached a P-value threshold of $<5x10^{-8}$ after filtering for MAF>1% and adjusting for covariates age, sex, single/doublet ICI therapy and presence of concurrent anti-cancer treatment. Variant annotation was performed using the gene annotation software tool ANNOVAR. Abbreviations: Chr- chromosome, CI- confidence interval (95%), Freq- frequency of effect allele, HR- hazard ratio, SNP- single nucleotide polymorphism, Pos- Pos- genomic position (build GRCh37)

5.2.3 Association analysis of HLA alleles

The *HLA* gene in the major histocompatibility complex (MHC) locus has been associated with multiple autoimmune diseases, which has similar clinical manifestations as irAEs[60]. Hence, I was interested in determining whether any *HLA* markers were linked with irAE development. Despite a large number of variants in the *HLA* region being assigned to the Illumina arrays used for genotyping our samples, only an average of up to approximately 45% of the known *HLA* alleles are genotyped[246]. To improve our genomic coverage of *HLA* alleles in our analysis, I imputed *HLA* alleles and coding SNPs using the T1DG reference panel and SNP2HLA as described in methods Section 2.15. I performed association analysis to determine whether classical *HLA* alleles and markers conferring changes to the amino acid sequences of *HLA* genes were associated with serious irAE. A total of 8961 *HLA* markers were analysed. **Table 5.17** shows the top variants within the chromosome 6 MHC region associated with serious irAE of any type. Two variants rs9295990 and rs2516472 map to an intergenic region between MHC class 1 chain-related protein A (MICA) and long intergenic non-protein coding RNA gene 1149 (LINC01149). MICA is known to be an inhibitory factor of cancer cell killing in the cancer immunity cycle[27]. Cancer cells shed MICA/B on their surface to evade immune surveillance and avoid anti-tumour responses by natural killer cells[247]. *MICA* gene has also linked with immune-mediated diseases including psoriatic arthritis and spondylitis[248].

Table 5.17 Top genetic variants within the chromosome 6 MHC region that are associated with the development of serious irAE

SNP	A1	A2	FREQ	OR	Lower CI	Upper CI	P value	Consequence	Mapped gene/nearest genes
rs6904669	G	A	0.43	1.59	1.19	2.11	0.002	intergenic	HCG27(dist=21051),HLA-C(dist=43730)
rs9295990	G	T	0.07	2.16	1.29	3.61	0.003	intergenic	MICA(dist=2927),LINC01149(dist=23425)
rs2516472	C	T	0.07	2.16	1.29	3.61	0.003	intergenic	MICA(dist=11608),LINC01149(dist=14744)
rs2596460	C	T	0.07	2.16	1.29	3.61	0.003	intergenic	LINC01149(dist=2760),HCP5(dist=13447)
rs2523691	T	C	0.07	2.16	1.29	3.61	0.003	intergenic	LINC01149(dist=5937),HCP5(dist=10270)
rs2596480	A	G	0.07	2.16	1.29	3.61	0.003	intergenic	LINC01149(dist=11235),HCP5(dist=4972)
AA_A_334_30021017_V	P	A	0.85	0.57	0.39	0.84	0.005	exonic	HLA-A
rs2256539	C	T	0.38	1.50	1.13	1.98	0.005	intergenic	HLA-A(dist=24079),HCG9(dist=5152)
AA_A_299_30020328_T	P	A	0.85	0.58	0.39	0.85	0.005	exonic	HLA-A

Summary statistics of amino acid polymorphisms, SNPs and HLA alleles within the MHC region associated with any serious irAE reaching a threshold of P value of <0.005. Imputation of the SNPs was performed using SNP2HLA and association testing using logistic regression analysis was performed using PLINKv1.9 adjusting for covariates age, sex, single/doublet ICI therapy and presence of concurrent anti-cancer treatment. A1 refers to the tested allele. Abbreviations: CI- confidence interval (95%), OR- odds ratio, irAE- immune-related adverse events, MHC- major histocompatibility complex, SNP- single nucleotide polymorphisms

5.2.4 Replication of previously reported SNPs associated with irAE

I performed a literature review to identify germline genetic variants previously reported in the other studies to be associated with irAE (Chapter 1, Section 1.4.6.1)[249]. I assessed the performance of these variants in predicting risk of serious irAE of any type in the discovery GWAS. **Table 5.18** shows the summary statistics for these SNPs in the discovery phase. Three SNPs, rs16906115, rs246079 and rs10964859 identified from two studies were associated with any serious irAE at nominal significance (P=0.049, 0.016 and 0.017 respectively). These SNPs mapped to genes *IL-7, UNG,* and *IFNW1* respectively.

Table 5.18 Summary statistics of previously reported SNPs in their association with any serious irAE in our discovery study cohort

Study	Sample size and population	Genetic analysis approach	Top SNP	OR	Lower CI	Upper CI	P value	Mapped Gene
Groha **[82]**	**1751 pan-cancer**	**All tagging SNPs**	**rs16906115[a]**	**1.66**	**1.01**	**2.73**	**0.049**	**IL-7**
Montaudie [127]	57 melanoma	All tagging SNPs	rs782637	1.15	0.86	1.52	0.344	PNPT1
Bins [118]	322 NSCLC	Candidate Study	rs2069705	1.13	0.85	1.51	0.394	IFNG
Abdel Wahab [117]	89 melanoma	All tagging SNPs	rs11711517	1.28	0.91	1.78	0.148	LOC105377125
Refae [122]	**94 pan-cancer**	**Candidate study**	**rs246079[a]**	**0.71**	**0.53**	**0.94**	**0.016**	**UNG**
			rs10964859[a]	**1.43**	**1.06**	**1.91**	**0.017**	**IFNW1**
Kirchhoff [126]	69 melanoma	All tagging SNPs	rs504963	1.14	0.87	1.50	0.337	FUT2
Marschner [129]	167 pan cancer	Candidate study	rs2910164	0.83	0.61	1.13	0.241	MIR146A
Queirolo [119]	173 melanoma	Candidate study	rs4553808	1.31	0.90	1.91	0.157	CTLA-4
Luo [125]	729 NSCLC	All tagging SNPs	rs9268543	1.37	0.98	1.91	0.066	HLA
Magis [121]	163 melanoma	HLA specific approach	HLA_DRB1_0401[b]	1.24	0.80	1.91	0.332	HLA
			HLA_DRB1_0301[b]	0.95	0.64	1.40	0.794	HLA

Association analysis results of the top SNPs reported from 10 studies with any serious irAE in our cohort are shown. Abbreviations: CI- confidence interval (95%), irAE- immune-related adverse event, NSCLC- non-small-cell lung cancer, OR- odds ratio [a] SNPs associated with any serious irAE from our GWAS discovery cohort at P<0.05 [b] Summary statistics of the HLA SNPs are reported from SNP2HLA association analysis for classical HLA alleles

5.2.5 Polygenic risk score analysis

GWAS have been useful in identifying individual SNPs that are associated with complex traits. However, these SNPs typically only have small effect sizes and correspond to small proportion of truly associated variants, hence limiting their predictive ability[87]. Most complex traits have been found by genetic studies to be contributed by more than one gene normally mapped by common variants of small effect sizes and rare risk variants of larger effect sizes. Although GWAS have helped us identified genetic loci that provide us insights into the mechanistic pathways of the disease, the ability to use these individual SNPs to predict disease is more limited[250]. Due to this limitation, polygenic risk scores (PRS) aggregate the effects of multiple variants into a single score that can help predict an individual's genetic predisposition to the disease[250]. The use of PRS in predicting the risk of developing irAE is therefore useful given that the irAEs are likely complex phenotypic traits that follow a non-Mendelian inheritance and influenced by more than one genetic marker.

5.2.5.1 PRS analysis based on SNPs associated with autoimmune disease

I extracted SNPs associated with 12 autoimmune diseases (AD) that reached GWAS significance ($P<5x10^{-8}$) from the GWAS catalogue. 58 SNPs were associated with >1 AD. PRS were calculated by measuring the sum of the risk alleles for that each individual carries weight by their effect size determined by a GWAS studying the phenotype of interest[87].The number of risk alleles per variant is weighted by the effect size from the largest European study reporting the genome-wide association with an AD.

Table 5.19 shows the ability of PRSs consisting of SNPs associated with the listed ADs to predict both serious irAE of any type and organ-specific irAEs. Apart from the any type serious irAE analysis, organ-specific PRS were only tested for association with ≥grade 3 irAE in the same organ. The PRS analysis consisting of 104 thyroid disease-associated SNPs was significantly

associated with any serious irAE (P=0.04), but counterintuitively with lower scores in those with toxicity than those without. A PRS containing all SNPs associated with ≥1 ADs (n=508 SNPs) was suggestive but not significantly (P=0.09) associated with serious irAEs; again scores were lower in cases than controls. In PRS analysis of organ-specific toxicities, the type 1 diabetes PRS was significantly associated with immune-related hyperglycaemia events (P=0.01) despite the small number of patients in our study with this rare irAE (N=7). For this PRS, the cases did have higher scores than the controls. **Figure 5.11 a) and b)** shows the distributions of the PRSs in controls and ≥grade 3 irAE cases using all AD PRS and the thyroid disease PRS. **Figure 5.11 c)** shows the distributions of the T1DM PRS in cases with and without immune-related hyperglycaemia/diabetic events.

Table 5.19 Summary statistics from PRS analysis of top SNPS correlated with different AD traits in predicting any serious irAE and associated ≥grade 3 end-organ toxicities.

Toxicity	AD trait	P value of model fit	Regression Coefficient	SE	No. of SNPs included in PRS
Any serious irAE	Addison's disease	0.38	-2.04	2.31	8
	AS	0.65	-1.53	3.32	37
	Coeliac	0.37	5.77	6.47	36
	Crohn's disease	0.50	13.37	19.87	12
	T1DM	0.16	9.88	6.95	46
	MG	0.83	0.29	1.41	6
	Psoriasis	0.52	-7.09	11.14	67
	RA	0.32	17.83	17.95	115
	SJS	**0.04**	**-2.00**	**0.98**	**3**
	SLE	0.06	-24.80	13.01	115
	AI thyroid	**0.04**	**-21.35**	**10.61**	**104**
	UC	0.73	-4.62	13.25	67
	Vitiligo	0.38	-6.54	7.43	64
	All AD	0.09	-37.77	22.60	508
GI (≥Gr 3)	IBD	0.93	-2.57	28.69	173
	UC	0.89	2.48	17.49	67
	Crohn's disease	0.39	-22.51	26.13	126
	Coeliac	0.58	4.73	8.59	36
Hyperglycaemia/DM (≥Gr2)	**T1DM**	**0.01**	**71.18**	**28.30**	**46**
Thyroid (any Gr)	AI thyroid	0.42	11.22	13.94	104

Skin (≥Gr 3)	SJS	0.19	-3.17	2.44	3
	Vitiligo	0.98	0.46	22.09	52
	Psoriasis	0.26	-35.91	31.68	67
Musculoskeletal (≥Gr 3)	RA	0.35	42.79	45.96	115
Endocrine (≥Gr 3)	Addison's disease	0.71	-1.60	4.36	8
Neurological (≥Gr 3)	MG	0.42	-4.53	5.65	6

Variants included in the individual AD PRS were identified from the GWAS catalogue from March-April 2021 and reached GWAS significance ($P \leq 5x10^{-8}$). The AD PRS highlighted in bold were associated with the relevant irAE at P<0.05. The 'All AD' trait comprises all SNPs associated with the listed AD traits. If a SNP was associated with >1 AD, the summary statistics with the most significant P value were used to weigh the SNP. Abbreviations: AD- autoimmune disease, AI thyroid- autoimmune thyroid disorders, AS- ankylosing spondylitis, GI- gastrointestinal, Gr- grade, GWAS- genome-wide association study, IBD- inflammatory bowel disease, irAE- immune-related adverse events, MG- myasthenia gravis, RA- rheumatoid arthritis, SE- standard error, SNP- single nucleotide polymorphisms, SJS- Steven Johnson's Syndrome, SLE- systemic lupus erythematosus, T1DM- type 1 diabetes mellitus, UC- ulcerative colitis

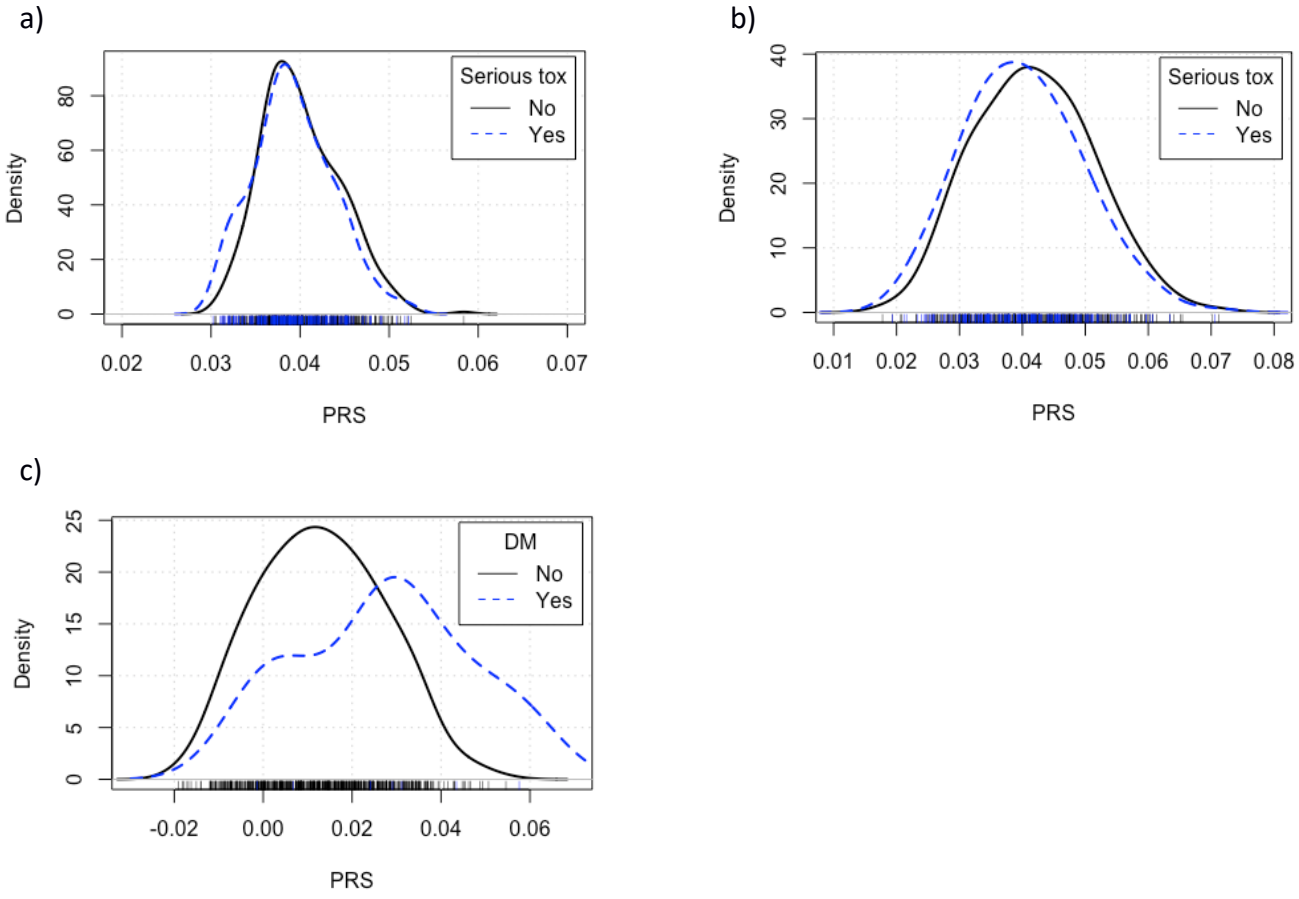


Figure 5.11 PRS prediction for irAE

Density curves showing the distribution of PRS stratified for presence and absence of the associated irAE in the a) prediction of serious irAE from all autoimmune SNPS (N=508), b) prediction of serious irAE from SNPs correlated with autoimmune thyroid disease (N=104) and c) prediction of diabetes from SNPS correlated with type 1 diabetes (N=46). Abbreviations: irAE- immune-related adverse event, PRS- polygenic risk score

5.2.5.2 PRS analysis based on SNPs associated with immuno-modulatory traits

Immuno-modulatory traits that are part of the cancer immunity cycle (Chapter 1: Section 1.4.4.1) include CD27, CD40/CD40 ligand, ICAM1, IFNα, IFNγ, VEGF, IL12, IL13, IL1, IL2, IL4, IL10, TNFα and MHC class 1 polypeptide-related sequence A and B (MICA/MICB) expression levels. SNPs associated with these traits were extracted from the GWAS catalogue and only those that reached GWAS significance were retained (n=83). The only PRS which was made from at least 20 SNPs was the PRS for MICA expression levels (n=27). After LD clumping and removal of ambiguous SNPs, only 12 SNPs remained for inclusion in the MICA PRS. This PRS was not significantly associated with ICI induced serious irAE (P=0.95).

5.2.5.3 PRS analysis based on top SNPs associated with GWAS analysis

In some cases, a PRS that includes more than just the top SNPs associated with a trait outperforms a PRS containing only GWAS significant SNPs[251]. I attempted to use LDpred2 to identify common variants that could be included in a PRS to predict serious irAE[182]. However, the SNP heritability score generated from the LD regression analysis was very small (<0.01). This observation indicates that the number of common variants that can jointly explain the differences in cases and controls is minimal, suggesting the weak predictive ability of PRS generated using this approach. I therefore only tested the performance of a PRS containing the top SNPs associated with serious ICI toxicity at $P<5x10^{-6}$ at identifying cases with serious irAEs. **Figure 5.12** shows the AUC curve, sensitivity analysis, and density curves of these top SNPs in predicting any serious irAE. The top eight independent SNPs associated with any serious irAE achieved an AUC of 77.8% with a sensitivity of 74.2% and specificity of 70.29% in predicting the presence of any serious irAE at a threshold of -0.09. This analysis is biased

given the markers used were identified in this cohort, and the performance of these eight SNPs as a PRS needs to be evaluated in an external independent cohort.

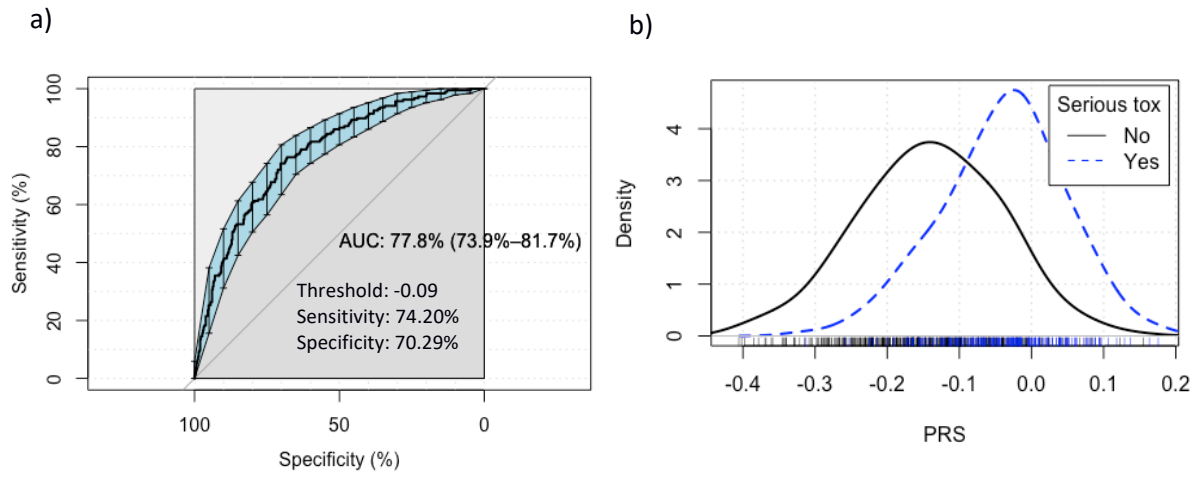


Figure 5.12 AUC-ROC curve and density plot of the top 8 SNPs associated with serious irAE in the prediction of irAE.

a) ROC analysis results for the top 8 SNPs associated with serious irAE listed in Table 5.3 in predicting the presence of any serious irAE. b) Density curves showing the distribution of PRS stratified for the presence and absence of any serious irAE from these SNPs in the prediction of any serious irAE.

These eight SNPs were therefore assessed for their performance in predicting any serious irAE in the validation cohort. As shown in **Figure 5.13**, the SNPs achieved an AUC of 51.4% with a sensitivity of 69.9% and specificity of 39.4% in predicting the presence of any serious irAE at a threshold of -0.13.

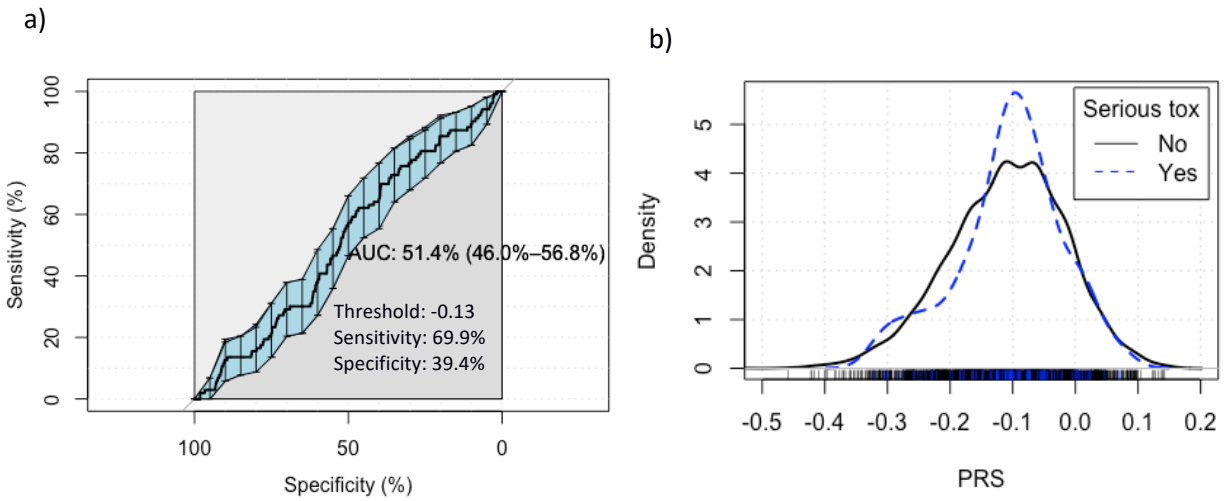


Figure 5.13 AUC-ROC curve and density plot of the top 8 SNPs associated with serious irAE in the prediction of irAE in the validation cohort

a) ROC analysis results for the top 8 SNPs associated with serious irAE listed in Table 5.3 in predicting the presence of any serious irAE in an independent validation cohort. b) Density curves showing the distribution of PRS stratified for the presence and absence of any serious irAE from these SNPs in the validation cohort.

5.2.6 Validation of top SNPs from GWAS

5.2.6.1 In silico evaluation in an external GWAS cohort

Given the modest size of the discovery cohort used to identify markers associated with irAEs, we wanted to identify a validation cohort. We collaborated with Sascha Gusev (Harvard University) who performed an in silico evaluation of the top 80 SNPs in his GWAS cohort from ~2000 patients, of which most had anti-PD1/PD-L1 therapy and the most common cancer diagnosis was NSCLC patients[134]. Only 23/80 of my top SNPs were available from their data where genotypes had been derived from off-target reads from panel sequencing of tumour samples using the OncoPanel at Dana Faber. Another difference to our data was the way the markers were analysed. They tested SNPs using a competing risk-based Cox proportional regression model with information on both time to toxicity and survival time.

Table 5.20 shows the performance of the SNPs from my GWASs of either organ-specific or any serious irAE in the Gusev validation cohort. The associated toxicity phenotype analysed in the Gusev cohort was the presence of ≥grade 2/3 irAE. Two SNPs from our discovery GWAS validated at P<0.05 in the validation cohort. rs4072644 associated with ≥grade 3 immune-related hepatitis in our GWAS ($P=3.2x10^{-6}$) was associated with the development of any ≥grade 2 or ≥grade 3 irAE in the Gusev cohort at a P-value of 0.035 and 0.049 respectively. GTEX expression quantitative trait loci (eQTL) data shows that rs4072644 is a significant eQTL to the *SLC1A3* gene with high expression in the thyroid ($P=4.2x10^{-6}$)[173]. The SNP has also been associated with immune-related traits including CD127 on CD28+ CD45RA+ CD8+ T cells ($P=4x10^{-4}$), interleukin 5 expression and monocyte counts(both $P=5x10^{-4}$)[252–254].

A second SNP rs55738551 associated with ≥grade 3 musculoskeletal events ($P=4.7x10^{-6}$) in our GWAS was associated with any ≥grade 2 irAE in the Gusev validation cohort with P=0.03. However, it had a low INFO score of 0.65 in our discovery cohort. We noted that the control MAF was quite different to the frequency reported in gnomAD[193], raising further concerns over the accuracy of the imputation data. We identified a proxy rs34511239 (r^2 0.99) for rs55738551 and looked this up in our discovery association data. The control MAF for this marker was consistent with gnomAD, and whilst the proxy was also associated with musculoskeletal events ($P=4.7x10^{-3}$), it did not reach our P-value cut-off.

Table 5.20 Summary statistics of the top SNPs identified from GWAS of irAE in our discovery cohort and a US validation cohort

Discovery								Validation (≥grade 2 irAE)		Validation (≥grade 3 irAE)		
SNP ID	**rsID**	**Toxicity**	**Grading (controls vs cases)**	**OR**	**P value**	**MAF**	**Risk allele**	**OR**	**P value**	**OR**	**P value**	**Mapped genes**
10:14755798_G_A	rs2609822	Resp	01v234	0.21	3.34E-06	0.59	A	1.00	9.6E-01	1.04	7.4E-01	LINC00929(dist=47985),GABRB3(dist=362525)
12:10222786_T_C	rs17734799	MSK	012v34	7.73	1.18E-06	0.17	C	1.05	7.6E-01	1.19	3.3E-01	CLEC1A
12:118039445_C_G	rs7960857	Endo	012v34	0.24	4.14E-06	0.81	G	1.07	5.2E-01	1.10	4.0E-01	KSR2
13:76176003_T_G	rs2328960	Skin	01v234	4.34	3.91E-06	0.85	G	0.91	4.3E-01	0.96	7.5E-01	UCHL3
13:111727635_G_T	**rs55738551**	**MSK**	**012v34**	**9.98**	**4.66E-06**	**0.08**	**T**	**1.24**	**3.3E-02**	**0.99**	**9.1E-01**	**LINC00431(dist=84686),LINC00368(dist=20550)**
15:31076481_C_T	rs12904674	Skin	01v234	0.19	6.40E-07	0.15	T	1.03	8.3E-01	1.23	1.7E-01	LOC100288637(dist=11272),HERC2P10(dist=33758)
19:53088882_C_G	rs35429660	Any ≥grade 3	012v34	0.45	5.13E-07	0.35	G	1.06	6.0E-01	0.91	4.7E-01	ZNF701
19:53088935_G_A	rs35764467	OST score	categorical	0.14	4.19E-08	0.38	A	1.06	6.0E-01	0.93	5.2E-01	ZNF701
2:5229400_C_A	rs6718493	Endo	012v34	5.73	2.95E-06	0.70	A	0.97	7.8E-01	1.01	9.4E-01	LINC01249(dist=525588),LINC01248(dist=544873)
20:50197080_T_C	rs8121883	Endo	012v34	0.15	3.98E-06	0.27	C	1.10	4.4E-01	1.11	4.3E-01	NFATC2(dist=17710),ATP9A(dist=16234)
21:29066227_G_A	rs57688665	Resp	012v34	32.08	2.29E-06	0.09	A	0.78	1.9E-01	1.21	4.4E-01	ADAMTS5(dist=726788),LINC00113(dist=28471)
4:14984760_C_T	rs4334749	Skin	01v234	2.46	4.08E-06	0.63	T	1.03	7.8E-01	1.05	6.3E-01	CPEB2-DT
4:31219796_C_T	rs6821356	OST score	categorical	0.47	4.73E-08	0.35	T	0.98	8.5E-01	0.93	5.8E-01	LINC02497(dist=6500),LINC02501(dist=288370)
4:40507858_G_A	rs36196909	Resp	012v34	19.71	4.50E-06	0.15	A	1.18	4.0E-01	1.14	5.5E-01	RBM47

5:1778380_T_C	rs11957851	MSK	012v34	7.93	2.46E-06	0.51	C	0.85	1.7E-01	0.90	4.0E-01	MIR4277(dist=69397),MRPL36(dist=20119)
5:36522484_T_C	**rs4072644**	**Hepatitis**	**012v34**	**0.22**	**3.23E-06**	**0.22**	**C**	**0.78**	**3.5E-02**	**0.77**	**5.0E-02**	**RANBP3L(dist=220473),SLC1A3(dist=83973)**
5:163065926_T_C	rs12657662	Endo	012v34	0.20	3.62E-06	0.34	C	1.05	6.4E-01	1.07	5.9E-01	MAT2B(dist=119567),LINC02143(dist=809502)
6:68050943_G_A	rs35938141	Endo	012v34	7.57	1.87E-06	0.06	A	0.86	4.7E-01	1.05	8.3E-01	SLC25A51P1(dist=1551567),LOC102723883(dist=714300)
6:123627922_T_C	rs9285450	Endo	012v34	0.25	8.40E-07	0.81	C	0.90	5.1E-01	0.81	2.5E-01	TRDN
7:145121057_C_T	rs6974853	Hepatitis	012v34	4.61	1.06E-06	0.10	T	0.94	7.1E-01	0.90	5.9E-01	TPK1(dist=587911),CNTNAP2(dist=692396)
8:31488655_C_T	rs4457297	Hepatitis	012v34	0.32	3.91E-06	0.44	T	1.07	5.5E-01	1.11	4.3E-01	WRN(dist=455268),NRG1(dist=8613)
9:789299_G_T	rs4742394	Resp	01v234	23.01	8.21E-07	0.73	T	1.14	2.2E-01	0.98	8.5E-01	FAM107B
9:86097564_A_C	rs35580298	Resp	012v34	0.00	7.46E-07	0.52	C	1.05	5.6E-01	0.90	3.3E-01	FRMD3

23 SNPs associated with any or organ-specific irAE were tested for their association with risk of any ≥grade 2 and ≥grade 3 irAE in a validation cohort of ~2000 patients treated with ICI[134]. SNPs highlighted in bold were significantly associated (P<0.05) with ≥grade 3 irAE in the validation cohort. Abbreviations: GWAS – genome-wide association study, Endo- endocrine, irAE- immune-related adverse event, MAF- minor allele frequency (risk allele), MSK- musculoskeletal, OR- odds ratio, Resp- respiratory, SNP- single nucleotide polymorphism

5.2.6.2 Validation in an independent cohort from ICI GENETICS

A validation cohort of 103 ICI GENETICS study samples was also formed to replicate the associations observed for the top SNPs identified from the discovery cohort. All 103 patients were recorded to have one or more serious irAE of any type. I collected their toxicity data and extracted their DNAs to be sent for genotyping using the same tagging SNP arrays as the discovery cohort. Subsequent genome-wide association analyses of serious irAE were performed by Claire Palles and a BMed Sci student in our group (Muthu Sivaswami). The control group consisted of 2497 individuals selected from UK BioBank diagnosed with melanoma, lung or kidney cancer. **Table 5.21** shows the performance of the top eight SNPs identified in the discovery cohort to be associated with any serious irAE at $P<5x10^{-6}$ in the validation cohort. Six SNPs had the same direction of effect in the validation phase, but with wide confidence intervals. None of the SNPs replicated at $P<0.05$ or reached GWAS significance in the meta-analysis of both discovery and validation cohorts.

Table 5.21 Summary statistics of the top markers associated with any serious irAE from the discovery GWAS in the discovery and validation cohorts and meta-analysis of both cohorts

rsID	Toxicity risk allele	Discovery OR (95% CI)	Discovery P-value	Validation OR (95% CI)	Validation P-value	Meta P-value
rs71399956	TTG	2.54 (1.79-3.61)	8.91E-08	0.82 (0.58-1.16)	0.26	3.80E-03
rs35429660	**C**	**2.21 (1.56-2.83)**	**5.13E-07**	**1.15 (0.85-1.56)**	**0.36**	**7.32E-05**
rs2472135	**A**	**2.10 (1.56-2.83)**	**5.83E-07**	**1.12 (0.85-1.50)**	**0.42**	**7.98E-05**
rs7197466	**T**	**2.08 (1.52-2.85)**	**2.19E-06**	**1.04 (0.76-1.42)**	**0.81**	**7.70E-04**
rs4527728	**T**	**2.21 (1.57-3.11)**	**2.68E-06**	**1.19 (0.88-1.62)**	**0.26**	**1.20E-04**
rs1389216733	**CA**	**6.34 (2.57-15.65)**	**3.07E-06**	**1.63 (0.76-3.51)**	**0.18**	**4.00E-04**
rs397892660	**CTT**	**2.48 (1.65-3.71)**	**3.59E-06**	**1.05 (0.72-1.53)**	**0.79**	**1.50E-03**
rs6864863	A	2.10 (1.52-2.91)	3.73E-06	0.76 (0.57-1.02)	0.07	0.11

Top SNPs predictive of any serious irAE at suggestive GWAS significance ($P<5x10^{-6}$) were tested for replication in the validation cohort, and a meta-analysis of both cohorts was conducted. Markers showing the same direction of effect in the validation phase are shown in bold. Abbreviations: CI: confidence interval, OR- odds ratio

I also evaluated the performance of the top SNPs associated with organ-specific ≥grade 3 irAE in the validation cohort, as shown in **Table 5.22** and **Table 5.23** for GI and hepatitis events respectively. Three SNPs associated with ≥grade 3 GI irAE, and seven SNPs associated with ≥grade 3 hepatitis had the same direction of effect in the validation phase. rs77941834 associated with hepatitis was replicated at nominal significance (P<0.05). The lack of validation with only one candidate SNP reaching nominal significance and wide confidence intervals render these results inconclusive.

Table 5.22 Summary statistics of the top markers associated with ≥grade 3 GI irAE from the discovery GWAS in the discovery and validation cohorts and meta-analysis of both cohorts

rsID	Toxicity risk allele	Discovery OR (95% CI)	Discovery P-value	Validation OR (95% CI)	Validation P-value	Meta P-value
rs112264014	**TTTG**	**10.92 (4.32-27.61)**	**5.13E-07**	**1.53 (0.37-6.41)**	**0.58**	**5.30E-06**
rs9552805	C	2.53 (1.72-3.72)	1.48E-06	0.63 (0.36-1.11)	0.10	2.87E-03
rs145650938	**T**	**6.43 (3.03-13.62)**	**2.08E-06**	**1.11 (0.42-2.93)**	**0.84**	**7.15E-05**
rs201085195	A	13.68 (4.85-38.59)	2.24E-06	0.16 (0.002-14.54)	0.29	3.57E-06
rs28481615	C	6.90 (3.08-15.43)	2.88E-06	NA[a]	0.01	2.64E-06
3:34716436_A_AGATins(41bases)	**A**	**28.39 (3.98-202.35)**	**3.44E-06**	**1.28 (0.39-4.15)**	**0.68**	**3.90E-02**
rs35651129	CA	2.98 (1.88-4.73)	4.53E-06	0.85 (0.42-1.70)	0.63	3.40E-04

Top SNPs predictive of ≥grade 3 GI irAE at suggestive GWAS significance ($P<5x10^{-6}$) in the discovery phase were tested for replication in the validation cohort, and a meta-analysis of both cohorts was conducted. Markers showing the same direction of effect in the validation phase are shown in bold. [a] rs28481615 was associated with large effect size (beta) of -1680.

Table 5.23 Summary statistics of the top markers associated with ≥grade 3 immune-related hepatitis GI irAE from the discovery GWAS in the discovery and validation cohorts and meta-analysis of both cohorts

rsID	Toxicity risk allele	Discovery OR (95% CI)	Discovery P-value	Validation OR (95% CI)	Validation P-value	Meta P-value
rs61770027	**G**	**8.09 (2.91-22.49)**	**7.56E-07**	**1.14 (0.52-2.49)**	**0.74**	**7.01E-03**
rs6974853	T	4.61 (2.53-8.41)	1.06E-06	0.31 (0.08-1.28)	0.05	7.93E-05
rs1157858353	**T**	**3.68 (2.14-6.33)**	**1.38E-06**	**1.06 (0.57-1.97)**	**0.85**	**2.56E-04**
rs4246254	**T**	**13.89 (4.69-41.16)**	**2.08E-06**	**1.22 (0.30-5.03)**	**0.79**	**8.44E-05**
rs550587077	**CT**	**18.29 (2.93-114.28)**	**2.61E-06**	**1.18 (0.48-2.94)**	**0.71**	**8.77E-02**
rs77941834	**T**	**5.17 (2.29-11.68)**	**2.74E-06**	**2.20 (0.94-5.15)**	**0.05**	**3.95E-05**
rs4072644	**T**	**4.63 (2.14-9.98)**	**3.23E-06**	**1.10 (0.59-2.06)**	**0.76**	**6.95E-03**
rs6509652	C	2.95 (1.81-4.82)	3.47E-06	0.74 (0.44-1.25)	0.27	1.68E-02
rs7571339	**C**	**2.94 (1.83-4.74)**	**3.69E-06**	**1.30 (0.78-2.19)**	**0.32**	**7.92E-05**
rs4457297	C	3.15 (1.87-3.15)	3.91E-06	0.67 (0.40-1.13)	0.13	4.77E-02
rs11659272	T	9.47 (3.70-24.24)	4.23E-06	0.95 (0.23-3.88)	0.94	1.14E-04
rs71368789	T	3.33 (1.98-5.62)	4.27E-06	0.66 (0.31-1.41)	0.26	1.82E-03

Top SNPs predictive of ≥grade 3 immune-related hepatitis at suggestive GWAS significance ($P<5x10^{-6}$) in the discovery phase were tested for replication in the validation cohort, and a meta-analysis of both cohorts was conducted. Markers showing the same direction of effect in the validation phase are shown in bold.

5.2.7 Transcriptome-wide association analysis (TWAS) for ≥grade 3 irAE

Findings from the discovery GWAS of serious irAE has helped identified genomic loci of interest linked with irAE. However, identifying the causal variant or gene driving the association can be challenging based on GWAS data alone due to certain limitations including linkage disequilibrium which can render the GWAS findings difficult to interpret[183]. Due to this, TWAS is a more robust method which uses expression reference panels from eQTL data

to identify gene-trait associations and help prioritise candidate genes for further functional analysis. I performed a TWAS based on summary level GWAS data using S-PrediXcan to identify genes whose expression was associated with serious irAE. The only gene that reached the significance threshold ($P<2.5x10^{-6}$) was *ZNF701* (effect size 4.93, $P= 5.1x10^{-7}$) expressed in the tibial arterial tissue. *ZNF701* is also expressed in 21 other tissues including colon, lung and thyroid with nominal significance ($P<0.05$), suggesting potential influence on toxicity from these organs.

5.2.8 Whole genome sequencing analysis of samples with multiple ≥grade 3 toxicities

Whole genome sequencing (WGS) to an average read depth of 30X was performed using germline DNA samples from eight patients with multiple ≥grade 3 irAE. These patients had two or more ≥grade 3 irAE and a OST score of ≥8. **Table 5.24** shows the toxicity profile of these patients.

Table 5.24 ≥Grade 3 irAE experienced by the 8 patients where WGS was performed on their samples

Sample no.	irAE toxicity									
	Skin	GI	Thy	Hepatitis	Resp	Renal	MSK	Blood	Ear	OST score
1	-	+	-	+	-	-	-	-	-	11
2	+	-	-	-	-	-	-	-	+	8
3	-	+	-	-	-	-	+	-	-	10
4	-	+	-	+	-	-	+	-	-	11
5	-	+	-	-	+	-	-	-	-	10
6	-	+	+	-	-	-	-	-	-	9
7	-	-	-	-	-	+	-	+	-	11
8	+	-	-	+	-	-	-	-	-	9

The types of ≥grade 3 events experienced by the patients are shown here, where '+' indicates the presence of a ≥grade 3 event and '-' indicates the absence of a ≥grade 3 event of the corresponding irAE type. The OST score refers to a cumulative score calculated for each patient sample based on the number of organs affected and the number of grade 3 events experienced with ICI therapy (scores: 1-3: low toxicity,4-7: moderate toxicity, 8-14: high toxicity group). Abbreviations: GI- gastrointestinal, irAE- immune-related adverse event, MSK- musculoskeletal, Resp- respiratory, Thy- thyroid

Using the GATK best practice guidelines described in Chapter 2.16, ~10 million variants passed QC, with ~8.6 million variants from autosomes available for annotation with VEP. Filters were applied to the annotated variants to identify novel rare variants (allele frequency <0.01 in gnomAD exome controls and genome controls (v3.1.2) that had possible loss of function, damaging/deleterious missense variants (from SIFT/Polyphen scores). Variants were also filtered to identify those mapped to a candidate panel of genes linked with immune-regulatory traits or from the GWAS of ≥grade 3 irAE (listed in Appendix 2). The number of variants identified from the filters applied is shown in **Table 5.25**.

Table 5.25 Types of VEP filter to the annotated variants

Variant consequences	**No. of variants**
Possible LOF (splice acceptor, splice donor, stop gained or frameshift)	277
Damaging/disruptive (SIFT ≤0.05 and Polyphen >0.908 scores)	481
Possible LOF and mapped to candidate genes	5
Damaging/disruptive and mapped to candidate genes	4

Filters were applied to VEP annotations to identify potential rare variants predictive of ≥grade 3 irAE that were not detected from imputation and directly genotyped data. Abbreviations: LOF- loss of function.

Other filters were then applied to identify rare variants annotated by ClinVar as pathogenic or likely pathogenic **(Table 5.26)**. One patient (sample no. 6), who had experienced immune-related thyroiditis was found to be a carrier of two frameshift variants, rs780952627 and rs772627768 mapped to *ATXN2*, which was associated with autoimmune thyroid disease in a GWAS[255]. One patient was a carrier of a likely pathogenic missense variant rs368114790C>T associated with diabetes mellitus in ClinVar. This variant maps to exon 10 of the *ABCC8* gene. Interestingly a top SNP (rs2237988) from our GWAS of ≥grade 3 endocrine irAE also maps to this gene. rs1800111G>C, a missense variant mapping to *CFTR,* was also carried by a patient (sample no. 5) who developed immune-related GI and pneumonitis toxicity. The variant is

reported to be of pathogenic significance for cystic fibrosis related disorders. Another variant, rs121918595C>T mapping to *RYR1,* is carried by a patient (sample no. 7) who developed severe ICI-induced renal and blood toxicity. rs121918595C>T is annotated in ClinVar as associated with toxicity to anaesthetic drugs.

Table 5.26 Top rare variants (allele frequencies <1%) from patients with multiple ≥grade 3 irAE

VEP filters	Chr	Pos	Ref allele	Alt allele	rsID	Consequences	Mapped gene	NFE AF in gnomAD	Associated trait of candidate gene/ClinVar	Sample no. of risk allele carrier	IrAE reported in risk allele carrier
Possible LOF and mapped to candidate genes	1	37931958	T	TA	rs143820864	Frameshift	INPP5B	2.32E-03	Autoimmune thyroid	5	GI, resp
	4	122649712	G	A	rs80146429	Splice donor	IL21-AS1	1.01E-02	Autoimmune thyroid	6	DM, thyroid, GI
	4	154237132	A	C	rs146298768	Stop gained	DCHS2	3.97E-03	Autoimmune thyroid	8	Skin, hepatitis
	12	111598992	GC	G	rs780952627	Frameshift	ATXN2	4.33E-04	Autoimmune thyroid	6	DM, thyroid, GI
	12	111598994	TGC	T	rs772627768	Frameshift	ATXN2	1.04E-04	Autoimmune thyroid	6	DM, thyroid, GI
Mapped to candidate genes and likely pathogenic	11	17442788	C	T	rs368114790	Missense	ABCC8	2.35E-04	Endocrine GWAS[a], Diabetes mellitus, Familial hyperinsulinism (LP)	3	GI, MSK
Possible LOF and pathogenic	1	152308647	G	T	rs147145234	stop gained	FLG	1.47E-04	Ichthyosis vulgaris (P)	4	GI, hepatitis, MSK
	4	145639572	C	T	rs104893851	stop gained	MMAA	3.53E-04	Inborn errors of metabolism-Methylmalonic aciduria (P)	8	Skin, hepatitis
Damaging variants and pathogenic/likely pathogenic	2	201078946	T	C	rs142609245	missense	NDUFB3	1.62E-03	Mitochondrial complex 1 deficiency (LP, P)	6	DM, thyroid, GI
	5	71604422	G	A	rs535519604	missense	MCCC2	1.47E-05	3-methylcrotonyl-CoA carboxylase 2 deficiency (LP)	6	DM, thyroid, GI
	5	111110280	G	A	rs116529882	missense	WDR36	1.69E-03	Open angle glaucoma (P)	4	GI, hepatitis, MSK

	6	51746853	C	A	rs148932323	missense	PKHD1	4.35E-03	Polycystic kidney disease (LP)	2	Skin, ear
	6	75131984	G	A	rs200487396	missense	COL12A1	1.76E-04	Congenital muscular dystrophy, Bethlem myopathy (US, P)	7	Renal, blood
	6	129349348	C	T	rs147077184	missense	LAMA2	3.90E-03	Congenital muscular dystrophy (US, B, P)	1, 5	GI, resp, hepatitis
	7	107694473	T	G	rs111033307	missense	SLC26A4	1.62E-04	Pendred syndrome, Autosomal recessive hearing loss (P)	1	GI, hepatitis
	7	117610521	G	C	rs1800111	missense	CFTR	2.43E-03	Hypertrypsinaenia, pancreatitis, cystic fibrosis (US, P)	5	GI, resp
	19	12896058	T	C	rs149120354	missense	GCDH	1.91E-04	Glutaric aciduria (LP, P)	3	GI, MSK
	19	38580094	C	T	rs121918595	missense	RYR1	not found	Response and toxicity to certain anaesthetic drugs (DR), malignant hyperthermia (P)	7	Renal, blood

VEP annotation filters were applied to identify possible loss-of-function, damaging (missense variants) or those mapped to candidate genes. Damaging/deleterious variants were identified using score thresholds of SIFT ≤0.05 and Polyphen >0.908. Further evaluation was performed to identify variants annotated by ClinVar to be pathogenic/likely pathogenic. All patient samples listed carried one risk allele of the corresponding risk variant. If the variant was mapped to the candidate gene, the associated trait of the gene is listed; otherwise the disease significance annotated by ClinVar is described; DR- drug response, LP- likely pathogenic, P- pathogenic, US- unknown significance. Abbreviations: DM- diabetes mellitus, GI- gastrointestinal, LOF- loss of function, MSK- musculoskeletal, NFE- non-Finnish Europeans, Resp- respiratory. [a]Mapped genes of top variants identified from our discovery GWAS

5.2.9 Correlation between toxicity and overall survival

The relationship between the development of serious irAE and survival rate was investigated in our initial cohort of 448 patients included in the GWAS of serious irAE who also had survival data available. The average follow-up duration for this cohort was only 1.8 years. 154 deaths were recorded at the time of analysis. **Figure 5.13** shows the unadjusted survival curves comparing patients who developed one or more serious irAE of any type versus those with no serious irAE. Patients who developed serious irAE (N=136) have better survival rates than patients with no serious irAE (N=312) with HR= 0.44, 95% CI= 0.30-0.66, P= 6.06×10^{-5}.

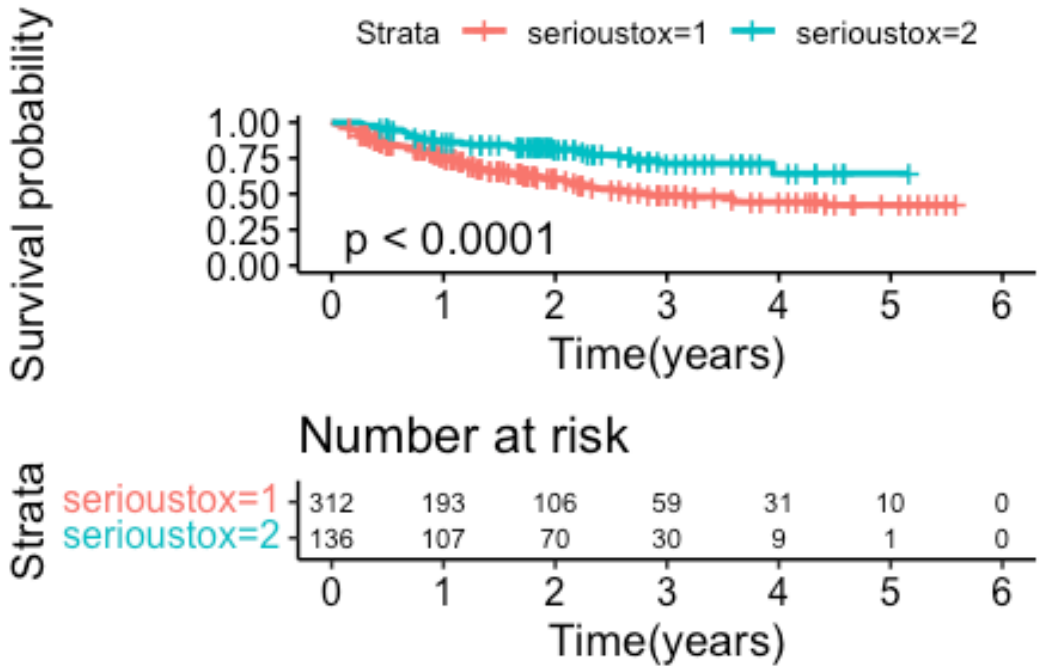


Figure 5.14 Unadjusted survival curves for patients with and without serious irAE (n=448)

Sex, age, type of cancer and cancer stage were included as covariates in an adjusted analysis, including patients where this information was available (N=237). A similar relationship was observed in the adjusted analysis where patients who experienced serious irAE (N=52) still had a significant survival benefit compared to patients without any serious irAE (N=185) with HR= 0.37, 95% CI= 0.17-0.82, P= 0.001.

5.3 Discussion

The ICI GENETICS study successfully recruited patients from the UK starting ICI treatment. As a result of the COVID-19 pandemic, recruitment was paused but thanks to collaborations set up with existing UK studies also aiming to investigate determinants of irAEs, we were able to collect data for a discovery cohort of 599 patients.

5.3.1 Toxicity profile of our discovery cohort

Most patients in the discovery cohort who experienced ≥grade 3 toxicities received doublet ICI treatment. This finding is in keeping with reports of higher rates of severe events in clinical trials of combination ICI therapies[31]. Cutaneous irAE were the most common any grade toxicity, mostly observed in anti-PD1 treated patients. This toxicity has been reported as common in this patient group[29]. GI irAEs were the most common ≥grade 3 event. This observation is consistent with previous studies, particularly in patients receiving doublet ICI therapy[31,256].

5.3.2 Markers of serious irAEs of any type

I did not identify any markers that reached the GWAS significance threshold ($P<5x10^{-8}$) in their relation to any serious irAE. Eight markers reached statistical associations of $P<5x10^{-6}$ but none reached statistical significance in the small validation cohort and further validation in a larger independent cohort is required.

One of these, rs35429660 **(Table 5.21)**, is a significant eQTL for *ZNF701*, the only gene to reach the significance threshold ($P=5.1x10^{-7}$) in the TWAS (Section 5.2.7). rs35429660 also reached

GWAS significance ($P=1.30x10^{-8}$) in its association with cumulative OST score, an alternative toxicity grading method for irAE, further strengthening its ability to predict risk of multiple severe irAE. Indeed, rs35429660 was also associated with more than one type of irAE with nominal significance, including hepatitis (P=0.01) and musculoskeletal toxicities (P=0.02). Hence, its association with any serious irAE in the initial GWAS does not seem to be driven by one organ-specific irAE alone. This variant has been reported to be linked with IL-17 levels, which has been identified as a potential biomarker for irAE and CD4 regulatory T cells involved in immunoregulation[153,252,257]. Interestingly, rs6509652 is another top variant associated with ≥grade 3 immune-related hepatitis that also maps to *ZNF701* and is in linkage disequilibrium with rs35429660 ($r^2=0.24$). The toxicity risk allele of these variants has mixed effects on *ZNF701* gene expression depending on the tissue site. The risk allele "*C*" of rs6509652 in particular, is associated with a reduced expression of *ZNF701* in CD4 T cells. Associations of these top toxicity-associated markers with CD4 T cells which have critical roles in the adaptive immune response, further highlight *ZNF701* as a strong candidate for investigating the biology of serious irAE.

The influence of *ZNF701* on the immune response is still not well defined. Among the immune cell types, the gene is predicted to be expressed in granulocytes but with low specificity[258]. *ZNF701* appears to have the highest expression in neutrophils and is expressed in CD4 and CD8 T cells to a lesser degree[258]. Considering the above findings, the proposed pathological mechanistic effect of this gene in irAE development could be explained via two processes: by controlling neutrophilic activity and subsequently mediating pro-inflammatory effects of irAE and/or via a gene knockdown effect causing an increase in CD4 T cells infiltration leading to disruption of immune homeostasis **(Figure 5.15)**.

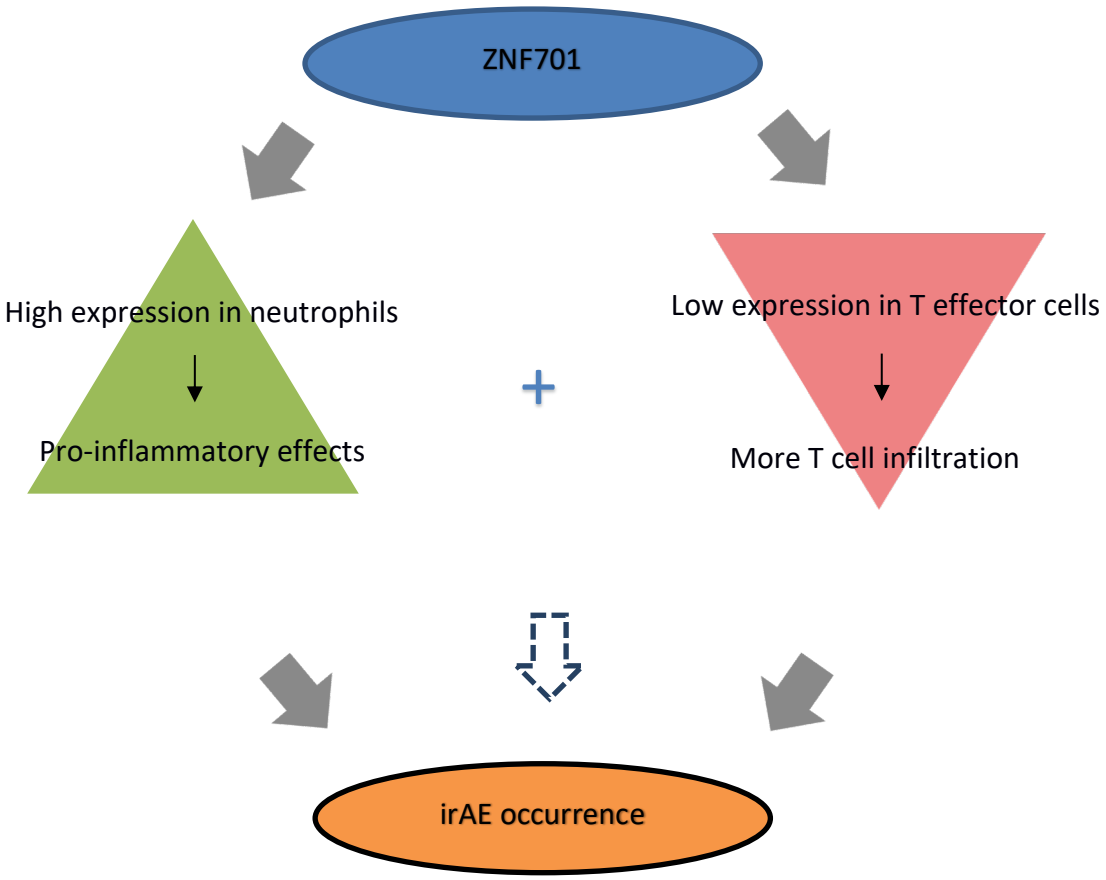


Figure 5.15 Proposed mechanisms of how ZNF701 expression influences irAE development
ZNF701 expression can potentially influence irAE development by mediating inflammation through neutrophilic activity and/or increasing immune cell infiltration through effector T cells at organs at risk.

I also tried identifying markers of any serious irAEs using a Cox-proportional hazards model incorporating time-to-onset of toxicity (Chapter 2: Section 2.3.4). I identified seven markers that reached GWAS significance, but we did not take this forward to validation as all markers had a MAF of 1% and would require a large validation cohort. The timing of toxicity was also only available for a lower number of patients in the discovery phase (N=542), limiting the power of this approach to identify variants associated with irAEs.

5.3.3 Markers of individual toxicities

I identified 75 markers associated with one or more individual irAE at $P<1x10^{-5}$. I attempted replication of these markers in a validation set from ICI GENETICS and UK Biobank. Whilst there were 103 cases in the validation phase with one or more ≥grade 3 irAEs, analyses of individual toxicities in this dataset were underpowered. GI and hepatitis toxicities were the most common ≥grade 3 irAEs in the validation phase with 30 and 35 cases respectively.

12 SNPs were associated with ≥grade 3 hepatitis and seven SNPs were associated with ≥grade 3 GI toxicities in the discovery phase at $P <1x10^{-6}$. Of the seven SNPs associated with GI toxicity in the discovery GWAS, three showed directions of effect in the validation cohort that were consistent with the results from the discovery cohort **(Table 5.22)**. rs112264014(ins TTTG) was associated with ≥grade 3 GI irAEs in the discovery GWAS (OR=10.92, $P=5.13x10^{-7}$). The SNP also showed the same direction of effect for the toxicity-associated allele in the validation phase with statistical support of $P=5.3x10^{-6}$ in the meta-analysis. rs112264014 maps to *MAST3*, which has been linked with inflammatory bowel disease (IBD)[259], known to have an autoimmune aetiology. rs112264014 is also a significant eQTL of *UPF1*, which regulates gut inflammation by modulating *TNFR1* expression in colonic epithelial cells[260]. *UPF1* and *TNFR1* have been identified as IBD susceptibility genes[260]. rs145650938C>T is another marker associated with GI events with a consistent direction of effect in the validation cohort. It is a significant eQTL for both *SNX6* and *KIAA0391*. *SNX6* is found to regulate PD-L1 checkpoint inhibitor expression, therefore, potentially determining the immune response to cancer cells and toxicity risk[261]. Genetic loci within *KIAA0391* have been identified to be associated with IBD in a transethnic meta-analysis[262]. The third variant 3:34716436_A_AGATins(41bases) which had a consistent effect in the validation cohort was near *PDCD6IP* and *LOC101928135*. Although *LOC101928135* is a RNA gene that is not been

well characterised, a study found that *PDCD6IP* expression in the rectal mucosae was downregulated in patients with ulcerative colitis who relapsed after discontinuation of their anti-TNF treatment compared to those who did not relapse[263]. This finding highlights the possible role of this variant in influencing immune-related colitis which present similarly to patients with IBD, but further functional work would be required to interrogate its interaction with this gene.

As shown in **Table 5.23**, one SNP (rs77941834) associated with ≥grade 3 hepatitis in the discovery phase was associated with same irAE at P=0.047 at the validation phase , the same direction of effect for the toxicity-associated allele and statistical support of P= $3.95x10^{-5}$ in the meta-analysis. rs4072644 was also associated with any ≥grade 2 and any ≥grade 3 toxicity in the Gusev replication dataset. Another SNP rs61770027 maps close to *IFI44,* which has been linked with multisystem inflammatory syndrome in children, viral hepatitis, immune cell infiltration in head and neck cancer and negative modulation of the innate immune response[264–267]. These seven SNPs , which have shown some evidence of replication may be worth considering in follow-up studies. Whilst the top irAE-associated SNPS identified so far have INFO scores of >0.4, it would be useful to check the accuracy of imputation for SNPs with an INFO of >0.4 but <0.8 using an orthogonal technique such as KASP genotyping.

The validation cohort had too few events of the rarer toxicities such as endocrine and musculoskeletal toxicities to provide any meaningful validation. Interestingly, two SNPs associated with endocrine and musculoskeletal toxicities from the discovery phase are also eQTLs for autoimmune genes and might be worth investigating in larger cohorts. rs2237988, associated with ≥grade 3 endocrine irAE (OR=4.21, P=$2.34x10^{-7}$) maps to *ABCC8* and is also a significant QTL for the expression of this gene in thyroid tissue. Strikingly, genetic mutations in *ABCC8,* which encodes sulfonylurea receptor 1, are associated with a subtype of maturity-

onset diabetes of the young (MODY) and intrafamilial diabetes[268][269][270]. rs1022831 associated with ≥grade 2 musculoskeletal irAE (P= $1.13x10^{-6}$), is a splicing QTL (sQTL) of *KANK1* expression in musculoskeletal tissue (P=$2.3x10^{-9}$), which is also a rheumatoid arthritis differently expressed gene[210,271].

Whilst I have identified several candidate SNPs that may affect the risk of specific irAEs, much larger sample sizes will be required to validate these. Subsequently, if these candidate markers were found to influence the risk of different irAE types at opposite direction of effects in the validation set, then implementing them in practice may be challenging.

There is also a lack of overlap between the top genetic markers associated with any serious irAE and organ-specific ≥grade 3 irAE. This observation may be because the mechanistic pathways of irAE vary between different organ systems. The absence of shared susceptibility loci between the different irAE types suggest possible variation in the immune profile of different organ types. Another explanation could also be related to other non-genetic factors driving irAE such as the microbiome.

5.3.4 Insights from gene-based tests

One of the top and largest ontology sets associated with any serious irAE risk include the set of genes upregulated in liver tissue upon *HNF1A* knockout (N=130 genes) **(Table 5.4)**. Mutations in *HNF1A* are known to cause pancreatic islet beta cell dysfunction leading to monogenic diabetes. Two of the major functional categories enriched in upregulated genes in *HNF1A* knockout islet cells include immune response and the TGF-β signalling pathway[272], highlighting the potential link of these genes with immune dysregulation leading to irAE.

HNF1A expression has also been associated with CD8+ T cell infiltration in cancer based on TCGA data[273], which is the likely pathophysiological mechanism of irAE.

Gene ontology analysis performed using the ≥grade 3 GI irAE GWAS summary statistics identified regulation of myeloid lymphocyte differentiation as an enriched gene set **(Table 5.7)**, suggesting potential combined roles of some GWAS significant associated markers in regulating inflammation, a key hallmark of irAE development. Interestingly, tissue expression analysis showed that the top genes correlated with GI toxicity have high expression in the GI tract organs **(Figure 5.8)**. This analysis supports the validity of the results for this irAE. It suggests there may be some genuine associations as expectedly, genes with high expression in the GI tract are functionally important and that markers influencing their expression could impact upon predisposition to GI irAE in the presence of ICI.

The top two gene-ontology sets identified from the GWAS of serious hepatitis are related to the biological process of tryptophan metabolism **(Table 5.10)**, which is known to play a role in inflammation through the kynurenine pathway. A reduction in free tryptophan due to increased IDO1 enzymatic action causes negative regulation of CD8+ T cells and inhibition of Th17 differentiation leading to immunosuppression[274]. Hence the reverse of this effect though influencing tryptophan metabolism could potentially determine the risk of immune-related hepatitis.

5.3.5 The use of polygenic risk scores to identify patients at risk of irAEs

I generated a PRS for the discovery phase participants using a set of 104 SNPs previously reported to be associated with autoimmune thyroid disease. This PRS was significantly associated with any serious irAE, indicating its potential in identifying patients at risk of serious

irAE. The PRS generated from 46 SNPs associated with T1DM was also significantly associated with hyperglycaemia events, suggesting the potential predictive ability of this score to detect patients at risk of immune-related diabetes, which often results in prolonged morbidity and reliance on long-term medication. Although there were only a small number of cases (N=7) in our patient cohort, none of the cases recorded any personal history of diabetes.

In addition to using PRSice2 for polygenic risk profiling, I used LDpred2 to evaluate the association statistics of all SNPs associated with any serious irAE (N=186 cases) to produce a PRS for irAE. However, the heritability score generated was low, indicating that the proportion of phenotypic variance explained by the common variants is minimal, thus compromising their predictive ability. This low score is likely due to the small sample size I used compared to the thousands of cases and controls LDpred was designed to analyse[182]. As datasets with irAEs and genetic data expand, designing an irAE PRS using this approach may be more feasible. Testing the PRSs I generated from the top irAE associated SNPs in independent samples would also be preferable. They have reasonable sensitivity and specificity in my discovery data, but this was expected as they were extracted from the same dataset that had identified them as toxicity-associated variants.

5.3.6 Whole genome sequencing

Results of the analyses reported so far are based on genotyping data from tagging SNP arrays, which would predominantly identify common variants of interest with MAF>1%. The ability to detect rare variants from these arrays is more limited. Therefore, I sent the samples of eight patients from the discovery cohort that had experienced multiple serious irAE for WGS to identify rare damaging or disruptive toxicity-associated variants. One patient (sample no. 6 from **Table 5.25**) was found to carry rare frameshift variants rs780952627del and

rs772627768del, which were mapped to *ATXN2.* Other variants mapping to *ATXN2* have also been identified from a GWAS meta-analyses of 18297 individuals for thyroid peroxidase antibodies positivity, which increases the risk of autoimmune thyroid disease[255]. Of interest, another patient who developed a serious respiratory irAE is found to carry a pathogenic missense variant rs1800111G>C (sample no. 5) mapping to *CFTR* which contributes to cystic fibrosis disorders, indicating that carriers of this variant could be susceptible to adverse respiratory conditions.

5.3.7 Alternative toxicity grading systems

NCI-CTCAE grades are used to score irAEs by most hospitals. I tested a scoring system ("OST score") that considered the number of affected organs and the severity of each organ-specific toxicity. The statistical support for rs35429660 was strongest in the analysis of the OST score ($P=1.30x10^{-8}$). rs35429660 was also second most significant SNPs associated with any serious irAE ($P=5.13x10^{-7}$), hence further supporting the "OST scores" as an alternative toxicity grading system to CTCAE. Given the quantitative nature of "OST scores", using this outcome variable in the association analysis is likely to have a higher power than using a binary outcome variable like presence/absence of ≥grade 3 irAE.

5.3.8 Relationship between ICI efficacy and toxicity

Several studies have assessed the association between the development of irAE and clinical outcomes of patients treated with ICI. A meta-analysis by Zhong et al involving 40 studies and 8641 patients of different cancer diagnoses treated with ICI who developed irAE had better objective response rates (ORR), overall survival (OS) and progression-free survival (PFS)

compared to those who had no irAE[275]. When stratified according to types of irAE, the study authors found that the development of skin, endocrine and pulmonary irAEs led to significantly improved PFS but this association was not observed in patients with pulmonary and hepatobiliary irAE[275]. The study also found that irAE occurrence was linked with favourable PFS in patients who received anti-PD1 therapy but not significantly evident among patients on anti-CTLA4 therapy[275]. Similar observations were also noted from a meta-analysis of 16520 melanoma patients where the non-thyroid endocrine and skin toxicities were significantly associated with better OS and PFS and patients on anti-PD1 had clearer clinical benefit than patients treated with anti-CTLA4 therapy[276]. The different clinical trajectory of the irAE types where cutaneous irAEs are generally easier to managed compared to lung irAEs could influence the difference in prognosis of patients who developed these irAEs. The lack of survival benefit seen with anti-CTLA-4 therapies could also be due to their association with higher morbidity and mortality requiring high doses of immunosuppression hence potentially compromising their clinical outcomes. Another meta-analysis by Hussaini et al evaluated 51 studies including patients with different cancer types on ICI improved clinical outcomes in patients who developed irAE regardless of type of cancer, irAE and ICI regime[277]. However, all three meta-analysis have found that high grade irAEs did not appear to have a clear survival benefit compared to patients with low grade irAE[275–277], which is in contrary to that observed in our dataset where patients who developed serious irAE had better survival rates. This could be due to the smaller sample size and the immature follow up survival data available for analysis in our cohort. There is also a high level of heterogeneity observed in the studies included in the meta-analyses with different study populations and design that has been noted as limitation of the analyses conducted, which could also contribute to the inconsistent observation noted with our analysis. To overcome the heterogeneity, more coherent data collection on response, survival and irAE stratified by

similar clinical factors would be required to better assess the relationship between irAE and clinical outcomes. It would also be clinically helpful to assess whether the lack of positive clinical benefit observed with certain irAE types and anti-CTLA4 therapies is similarly seen in our dataset once we have sufficient follow up data collected.

5.3.9 Strengths and limitations

The strengths of my data are the bespoke toxicity collection, where part of the ICI GENETICS study design includes a prospective recruitment cohort who had their toxicity data collected from their first cycle of ICI treatment. Another strength of my analysis was the high quality and high-resolution genotyping data I generated. We performed a full scan of the whole genome using tagging SNPs rather than relying on off-target reads from tumour sequencing, which enabled us to achieve better coverage of the whole genome. Being able to start validating our data in an independent cohort was also a strength of our analysis.

The main limitation, however, is the relatively small sample size compared to GWAS of other complex traits. Due to this, the analysis would only have been powered to detect common variants at large effect sizes. However, recent pharmacogenomics studies have often detected candidate variants with large effect sizes from smaller sample sets, suggesting that sample size may not be necessarily a crucial factor in GWAS when studying drug-related outcomes[85].

I attempted to use the Gusev validation cohort to validate the top toxicity-associated SNPs from the discovery phase. However, differences in genotyping methods and toxicity data coding and analysis between our studies preclude this. These differences also limit the ability to perform a meta or pooled analysis of our datasets to improve the power of our analysis.

Hence, I next validated our findings in an independent cohort of 103 cases that were genotyped using the same genotyping platform and had their toxicity data coded in a similar manner as our discovery cohort. However, this validation phase is potentially limited by contaminating risk alleles in the controls from UK BioBank as it uncertain whether a proportion of the controls may also have a serious irAE if they received ICI treatment. We tried to negate this limitation by having a large sample size for our control group. We were mindful that the validation cohort would only be helpful if there were variants of large effects to detect. The lack of replication in the validation cohort indeed suggests that the true effect sizes of many of the variants identified in the discovery phase are smaller.

5.3.10 Future directions

Overall findings from fine-mapping the top toxicity-associated variants have helped determine potential causal mechanisms of irAE. Some of the associations and pathways identified involve processes relating to immuno-regulation, auto-immunity and DNA repair. Future work will include designing functional experiments to test the downstream implications of variants detected from the discovery and validation phases and in the meta-analyses of both cohorts. Results of our datasets can contribute to a meta or pooled analysis with other studies of similar design and study outcomes to further improve the power of the GWAS in identifying potential genetic determinants of serious irAE.

Given the detection of several promising rare risk variants carried by patients with multiple serious irAE from the WGS data, I will next perform gene-burden testing to establish rare variants gene groups predictive of serious irAE.

The top genetic variants predictive of irAE can also contribute to a predictive model of toxicity risk alongside other biomarkers of toxicity as discussed in Chapter 1: Section 1.5.3. A practical model can guide decisions on personalised treatment options and safety monitoring of toxicities based on the patient's risk. Using this model in clinical practice will allow prompt detection of patients at high risk of serious irAE who may require close surveillance and early intervention to reduce the risk. It will also be helpful to determine in independent sample sets whether markers identified here can predict toxicity in real-world clinical practice.

The positive relationship between irAE development and survival outcomes from ICI therapy highlights the possibility of shared genetic loci contributing to both treatment outcomes. Further association analyses to identify germline variants predictive of survival and treatment efficacy would be of value when the survival data from our initial GWAS discovery cohort is mature. It would be interesting to see whether the top SNPs identified from our GWAS of serious irAE also predict improved survival and vice versa.

CHAPTER 6: DISCUSSION

6.1 Toxicity induced by chemotherapy or immunotherapy

Advances in cancer treatment have played a huge role in improving cancer survival with a 10-year survival rate of 50% for patients diagnosed with cancer in England and Wales[2]. Despite the considerable successes of these therapies, their associated toxicities can cause profound morbidity and, in some cases, potentially life-threatening. The impact of the toxicities experienced by patients on their quality of life is a significant challenge to consider in cancer survivorship as overall survival rates continue to improve. Some toxicities may result in life-long reliance on other medications and adverse effects on the mental health of patients who already have to cope with the detrimental effects of their cancer diagnosis. High costs to stretched healthcare systems can also occur from these adverse events. Achieving the right balance to obtain high treatment efficacy with minimal toxicities remains challenging in patients receiving anti-cancer therapy.

6.2 Role of germline genetic variation in predicting toxicity

The ability to identify patients at high risk of severe toxicity is appealing to provide a more personalised approach in the management of cancer patients. Variation in our host genome has been strongly hypothesised to be an important factor in determining a patient's risk of toxicity. The influence of inherited variation has been demonstrated in established pharmacogenomics research, where causal risk variants such as those mapping to the *DPYD* gene have been identified to be associated with serious fluoropyrimidine toxicity. Four *DPYD* variants are currently being tested in clinical practice before fluoropyrimidine chemotherapy administration to identify patients at high risk of severe toxicities and guide treatment dose modifications. Oral capecitabine is a prodrug of 5-FU that has also been widely used to treat

gastrointestinal and breast cancers. Although several studies have been conducted to identify potential genetic variants predisposing towards capecitabine toxicity, findings from these studies have been inconsistent and limited by small sample sizes, heterogeneity in sample populations and drug regimens. The lack of validation for the potential variants identified indicates the need for larger studies using a more standardised study population cohort and chemotherapy regimes.

Immune-related adverse events (irAE) induced by immune checkpoint inhibitors (ICI) have been observed to mimic autoimmunity and occur due to immune checkpoint blockade affecting immune homeostasis. Well-established associations between the host genome and immune-mediated traits further support the role of germline genetic factors as determinants of immunotherapy toxicity. Various methodology approaches have been used by studies to date in the identification of genetic factors predisposing towards treatment toxicity. These include using small-scale candidate gene panels, imputation from genotyping tagging SNP arrays, HLA typing and whole exome sequencing. Findings from these studies support the importance of germline genetics factors as key determinants of irAE occurrence. However, most studies are limited by small sample sizes and therefore only powered to detect genetic variants of large effect sizes. The studies were mainly retrospective with either limited or no validation of the variants identified. Larger population cohort genetic studies are essential to produce meaningful results and associations that can be translated into clinical practice. Together with other promising biomarkers of toxicity, the detection of potential germline genetic markers would help contribute towards a predictive model of toxicity risk.

6.3 Using GWAS to identify candidate genetic markers of toxicity

Genome-wide association studies (GWAS) were used as the main analytical method in identifying predictive genetic determinants of fluoropyrimidine and ICI-induced toxicity projects conducted in my PhD. GWAS have been widely and successfully used to identify associations of genetic loci with common complex disease traits and can help reveal potential molecular mechanisms underlying these traits. However, genome-wide analyses rely on large sample sizes to be adequately powered to detect variants with significant effect sizes, thus may not be as reliable for identifying associative variants for some of the rare toxicities. Examples of these include certain organ-specific serious irAE involving the respiratory, cardiac and neurological systems, which occurred in ~1% of patients treated with ICI in our discovery cohort. Early capecitabine-induced serious toxicities such as haematological and vomiting events are also uncommon affecting ~1% of patients from the QUASAR2 trial. The use of GWAS in identifying potential candidate SNPs that can help predict patients at risk of these events is therefore more limited.

Variants identified by the GWAS of both toxicity projects conducted are in non-coding or intergenic regions of the genome, consistent with the findings reported in GWAS of complex diseases. These variants have no direct effects on the coding sequences of the gene, but fine-mapping the lead variants has shown that many alter the expression levels of genes, perhaps because they are located within regulatory elements and affect transcription factor binding sites[278]. GWAS often identify multiple variants in linkage disequilibrium (LD) with one another and are inherited together during meiotic recombination[278]. LD between markers can make identifying the true causative variants at each locus difficult. I used additional conditional and joint analysis tools such as GCTA-COJO to identify independent loci by correcting for LD estimated from a reference sample and analysing the joint effects of the

SNPs[171]. Single-SNP association analysis from GWAS does not consider the total variation at a locus. As there could be multiple causal variants at a locus, I performed conditional analysis on the primary associated SNPs to detect secondary associated SNPs of interest. Following the identification of the top independent genetic variants from the genome-wide analyses performed, it is often still unclear which genes are regulated by these variants and in which cell types or physiological context. Colocalising GWAS signals with molecular phenotypes like gene expression quantitative trait loci (eQTLs) enabled us to identify genes regulated by the candidate causal variants. By using this approach to fine-map top independent SNPs linked with serious toxicity using eQTL data from the GTEX portal together with chromatin interaction mapping from Hi-C data, I was able to prioritise potential candidate genes that can help guide further functional work. This approach highlights that GWAS findings should be integrated with functional genomics datasets to achieve biological insights into the mechanisms of toxicity.

The next steps required to establish the functional role of the top disease-associated SNPs and work out how they lead to toxicity can be complex, requiring a prolonged period and high operational costs. It is therefore debatable as to whether this step in the research pipeline is necessary before using these SNPs as a screening tool, particularly if the variant has been validated successfully in large datasets. There are also currently more advanced in silico fine-mapping tools that can help predict the functional and structural consequences of variants. An example of this is the *DPYD*-specific variant classifier developed by Shrestha et al. which uses machine learning and in-vitro enzyme activity data from 156 missense *DPYD* variants to identify variants that can affect 5-FU catabolism and increase the risk of 5-FU toxicity in different ethnic populations[279].

As observed with complex disease traits, toxicity risk is polygenic, i.e. multiple genetic variants explain the risk. Combining analysis of multiple risk variants into a single score like a polygenic risk score (PRS)) to reflect each individual's risk may produce a more sensitive risk prediction. Polygenic risk profiling is increasingly being used following the discovery of tens to hundreds of genetic variants associated with a single trait or disease by GWASs. The accuracy of PRS depends on large sample sizes used to identify and validate multiple variants concerning a trait of interest. Only one marker has been previously published at genome-wide significance relating to ICI-induced irAEs[82]. We failed to identify any additional ones that reach this threshold with validation in a second cohort. A PRS to identify those at risk of ICI-induced irAEs is a future objective but we are not there yet. Collaboration between groups will be required to achieve the large sample sizes necessary to identify the full spectrum of variants influencing the risk of irAEs, which can be combined to generate a PRS with high sensitivity and specificity that can be translated to clinical practice.

6.4 Challenges in conducting studies of irAE

Studies to identify determinants of treatment-related irAE are lagging behind those researching determinants of treatment efficacy. Toxicity assessments are often secondary outcomes of large randomised clinical trials of cancer treatments. Standardised toxicity severity scores such as CTCAE have been used in trial and clinic settings to guide toxicity grading. CTCAE was initially developed before immunotherapy was introduced as a systemic cancer treatment modality. Although it has been effective in grading toxicities related to conventional chemotherapy, some toxicity definitions may not be as applicable in grading irAE severity. There are currently no established scoring systems for irAE; hence, the OST cumulative scoring system developed by the PAIR study team from King's College London was

designed to test the ability of an alternative system to grade and capture irAE more accurately. Outside clinical trials, there is variability in the extent of irAE reporting and recording in medical notes. This variation can lead to inconsistent and unreliable data collection and evaluation of toxicity events in the real-world clinical setting. We are also unaware of ICI toxicity data associated with samples collected by large projects such as the UK Biobank or 100,000 genomes project being made publicly available for analysis. This data is hard to extract from medical records at any scale. Differences in toxicity data collection between sites can also lead to challenges in collaborative research efforts which are often required to generate large enough datasets for uncommon toxicity events. Standardised, regular, and up-to-date recording of toxicity events should be encouraged as it will not only be beneficial when reviewing and assessing patients for toxicity in the clinic, but at the same time it will also help future research involving treatment-related toxicity to improve patient outcomes and cancer care.

6.5 Are genetic variants predictive of toxicity associated with better treatment efficacy?

Identification of predictive biomarkers of toxicity has clear benefits in reducing the risk and protecting the patient from treatment-related severe toxicities. However, growing evidence shows that patients who developed toxicities also responded better to their treatment.

Although the survival data of a proportion of patients in our discovery GWAS cohort of ICI toxicity is yet to mature and still being collected, I was able to analyse the survival data of 448 patients who had been followed up for an average of 1.8 years. Patients who developed any serious irAE had significantly better survival than those with no serious irAE. Positive

correlations between cancer treatment-related toxicity and response have also been reported in the literature. A systematic review by Teulings et al. of 27 studies involving patients receiving immunotherapy found that 3.4% of patients who experienced vitiligo had a significant (2-4 times) lower risk of disease progression or death[280]. The severity of toxicity could also be correlated with the degree of response. In a study of head and neck cancer patients treated with cetuximab, patients who developed severe grade 3 or higher skin rash had significantly better progression-free survival and overall survival than those with lower grade toxicity [281].

It is unclear whether the same biomarkers associated with higher risk of toxicity are also related to increased drug response and efficacy. Some studies assessing for genetic correlates of ICI efficacy have identified associated SNPs linked with immune-related genes and autoimmune diseases. A recent GWAS by Groha et al. identified a SNP near IL7 that reached GWAS significance in its association with all-grade irAE[82]. The SNP was not however associated with survival in patients treated with ICIs. Interestingly however, the SNP was associated with increased survival in melanoma patients who are mostly not treated with ICI as part of the Cancer Genome Atlas(TCGA) project[82]. Data from Taylor et al. suggests this SNP may be involved in melanoma progression[282]. One potential challenge would be the discovery of shared genetic susceptibility loci associated with both treatment efficacy and toxicity. Therefore, a patient's predicted risk of toxicity would need to be considered alongside information on markers of efficacy to maximise treatment benefit while reducing the risk of serious toxicities.

6.6 Future direction

The top SNPs identified from the meta-analysis of early capecitabine toxicity will require further functional work to gain biological insights into the mechanisms leading to early toxicity and assess their predictive ability in independent sample sets. Further genotyping will be needed for one variant identified in the COIN study dataset to be associated with early haematological toxicity at GWAS significance. This variant was not genotyped or imputed in the QUASAR2 dataset. Six variants reaching GWAS significance in the GWAS meta-analysis of cardiotoxicity but not present in all three studies analysed will also need extra genotyping. These top variants identified from the meta-analyses GWAS of early capecitabine serious toxicity and cardiotoxicity will potentially contribute to the genetic variant panel currently used in screening tests before administering 5-FU or capecitabine therapy. Considering these variants in the tests may help improve the sensitivity of the current test in identifying patients at risk of serious fluoropyrimidine-related toxicities and guide further dose modification and treatment decision-making.

The validation cohort for the ICI toxicity GWAS as discussed in Chapter 5 consisted of 103 cases with one or more serious irAE genotyped using the same tagging SNP array platform as the discovery samples. The cohort also included 2497 patients with cancer (matching the cancer types of the irAE cases) selected from UK BioBank as the controls. We recognise that the sample size of cases is still relatively small; thus we are looking to add 67 samples already genotyped from the Oxford group whose toxicity data are pending. We will also plan to obtain additional samples from other groups based in Cambridge and Imperial College London. If the top SNPs identified from the discovery GWAS have been validated in this independent cohort, further functional experiments to test the downstream implications of these variants can be planned. For example, rs35429660 is one of the top variants associated with risk of serious

irAE in our discovery cohort, of which its risk allele is found to influence the expression of *ZNF701*. A functional assay can be designed to measure and analyse *ZNF701* gene expression in isolated immune cells (e.g. B cells and T cells) from patients treated with ICI compared to healthy controls. Further gene knockdown or overexpression experiments can then be conducted to assess the effects of the gene in effector immune cells of patients who developed serious irAE compared to those with no or mild toxicity.

The discovery and validation datasets can also be combined for further association analyses to reach a larger sample size. Results from these datasets can contribute to a meta or pooled analysis with other studies with similar designs and study outcomes to further improve the power of the GWAS in identifying candidate genetic determinants of serious irAE. Although performing a meta-analysis with the large GWAS of irAE published by Groha et al. would have been desirable, a re-analysis of our data using a similar toxicity coding criteria and genotyping method to their study would be required.

Applying a predictive risk model of toxicity in clinical practice can help guide decision-making on cancer treatment options. This decision is particularly relevant when considering the need for curative intent adjuvant treatment following radical surgery, where most cases would have achieved the maximal benefit with surgery alone. One would most likely decide against adjuvant therapy if the patient was predicted to be at high risk of potentially fatal or life-changing toxicities and the likely treatment benefit is minimal. Toxicity can be long-lasting and associated with chronic morbidity and dependence on lifelong medication. We know from NCRI Living With and Beyond cancer groups that quality of life is vital to cancer survivors and needs to be protected when making treatment decisions. Balancing the pros and cons of treatment in patients predicted to have a low or moderate risk of serious toxicity from their

germline genotype would be more difficult, especially if they have a good chance of responding to a highly effective treatment.

Discoveries from GWAS of treatment-related toxicity can also help evaluate the potential mechanisms and pathways behind the development of these adverse events. Chemotherapy-induced toxicity is most likely associated with abnormalities in the proteins or molecules involved in the drug catabolism pathway. In contrast, immunotherapy toxicity has been hypothesised to be caused by an overactive immune system due to disruptions in immune homeostasis. The genetic markers detected to be correlated with toxicity would help identify the likely causative enzyme protein or immune modulator driving these events and define their potential mechanism of pathogenicity. These discoveries will help guide future functional research on prophylactic and therapeutic interventions and strategies in toxicity management.

The identified inherited genetic markers can be used alongside other biomarkers and form a predictive biomarker panel of toxicity. Biomarkers predictive of toxicity that have been studied to date include the measurement of pro-inflammatory cytokine and chemokines, composition of the gut microbiome, B cell changes and early diversification of T cells repertoire.

Genetic variants associated with irAEs should also be investigated as potential genetic determinants of ICI treatment efficacy, given the potential relationship between ICI toxicity and response. Using predictive markers or models of treatment efficacy and toxicity as a screening tool in practice will contribute towards personalised treatment to balance the risk of toxicity with the chance of treatment response for each patient. These predictive tools will help guide decision-making between the clinician and patient on their cancer management. A

conservative approach with stringent criteria has been the approach to implementing genetic testing before treatment with fluoropyrimidine-based treatments. A similar approach will be required for variants mandated for testing before ICI treatment to ensure that only variants whose impact can be interpreted are tested for and the results can be clear for both the clinicians and patients. It is only worth including clinically meaningful variants with good sensitivity and specificity. Interventions that can help prevent the development of serious toxicity for at-risk groups include the implementation of a surveillance protocol to monitor for toxicities, the use of supportive medication that can prophylactically prevent the worsening of toxicities as well as training and education of patients, carers and the wider health community on how to recognise and manage early signs of toxicity.

Findings from studies on predictive toxicity biomarkers can also guide research on personalised treatment strategies in managing irAE. Manipulation of the epigenetic pathways, targeted therapies towards B cells, autoantibodies targeting pro-inflammatory cytokines, and steroid-sparing interventions are some of the potential treatment strategies that will need further research to help reduce the severity of irAE in patients receiving ICI[15][155][283]. Pai et al. developed a modified version of the anti-CTLA-4 antibody, the main ICI agent used to treat metastatic melanoma[284]. They designed a dual variable domain immunoglobulin of anti-CTLA which concurrently depletes tumour infiltrating regulatory T cells while sparing tissue-resident T regs, leading to potent anti-tumour effects while reducing tissue-specific toxicity[284]. The need for continuing research in developing novel toxicity prevention and treatment strategies is essential as more patients are likely to be treated with ICIs as their treatment indications are growing rapidly.

6.7 Concluding remarks

My research project has identified potential genetic determinants of capecitabine chemotherapy and ICI-induced serious toxicities. Some of these variants have been fine-mapped to candidate genes that could explain the mechanisms behind developing these toxicities and guide future functional work, although independent validation is required. These genetic markers may help improve the ability of genetic panel tests to identify patients at risk of severe toxicities before administration of these therapies. Being able to detect patients at risk will help plan their further anti-cancer treatment and guide the necessary toxicity prevention and surveillance strategies accordingly. The toxicity profile determined by these markers of toxicity will need to be considered alongside other markers of efficacy to provide patients with the maximum treatment benefit and minimal serious toxicity. It is also important to consider the negative impact of these adverse events on the patient's day-to-day activities. The ability to predict patients at risk using these genetic biomarkers to effectively prevent and promptly treat unwanted irAE will help reduce morbidity and mortality rates, hospital admissions, healthcare cost burden and improve patients' overall quality of life.

CHAPTER 7: REFERENCES

1 Chin IS, Galavotti S, Yip KP, *et al.* Influence of clinical characteristics and anti-cancer therapy on outcomes from SARS-CoV-2 infection: a systematic review and meta-analysis of 5,678 cancer patients. *medRxiv* Published Online First: 2020. doi:10.1101/2020.12.15.20248195

2 Cancer Research UK. Cancer survival statistics. Cancer Res. UK. 2014.https://www.cancerresearchuk.org/health-professional/cancer-statistics/survival/all-cancers-combined#heading-Zero (accessed 22 Apr 2020).

3 CancerResearchUK-NHS England. Chemotherapy, Radiotherapy and Surgical Tumour Resections in England. https://www.cancerdata.nhs.uk/treatments (accessed 20 Sep 2023).

4 Meulendijks D, Cats A, Beijnen JH, *et al.* Improving safety of fluoropyrimidine chemotherapy by individualizing treatment based on dihydropyrimidine dehydrogenase activity- Ready for clinical practice? *Cancer Treat Rev* 2016;**50**:23–34. doi:10.1016/j.ctrv.2016.08.002

5 Diasio RB, Harris BE. Clinical Pharmacology of 5-Fluorouracil. *Clin Pharmacokinet* 1989;**16**:215–37. doi:10.2165/00003088-198916040-00002

6 Longley DB, Harkin DP, Johnston PG. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer* 2003;**3**:330–8. doi:10.1038/nrc1074

7 Nazki FH, Sameer AS, Ganaie BA. Folate: Metabolism, genes, polymorphisms and the associated diseases. *Gene* 2014;**533**:11–20. doi:https://doi.org/10.1016/j.gene.2013.09.063

8 Lam SW, Guchelaar HJ, Boven E. The role of pharmacogenetics in capecitabine efficacy and toxicity. *Cancer Treat Rev* 2016;**50**:9–22. doi:https://doi.org/10.1016/j.ctrv.2016.08.001

9 Morita T, Matsuzaki A, Kurokawa S, *et al.* Forced Expression of Cytidine Deaminase Confers Sensitivity to Capecitabine. *Oncology* 2003;**65**:267–74. doi:10.1159/000074480

10 Mercier C, Dupuis C, Blesius A, *et al.* Early severe toxicities after capecitabine intake: possible implication of a cytidine deaminase extensive metabolizer profile. *Cancer Chemother Pharmacol* 2008;**63**:1177. doi:10.1007/s00280-008-0889-1

11 Bonotto M, Bozza C, Di Loreto C, *et al.* Making Capecitabine Targeted Therapy for Breast Cancer: Which is the Role of Thymidine Phosphorylase? *Clin Breast Cancer* 2013;**13**:167–72. doi:https://doi.org/10.1016/j.clbc.2012.10.002

12 Rosmarin D, Palles C, Pagnamenta A, *et al.* 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* 2015;**64**:111 LP – 120. doi:10.1136/gutjnl-2013-306571

13 Topalian SL, Drake CG, Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* 2015;**27**:450–61. doi:10.1016/j.ccell.2015.03.001

14 Menon S, Shin S, Dy G. Advances in Cancer Immunotherapy in Solid Tumors. *Cancers (Basel)* 2016;**8**:106. doi:10.3390/cancers8120106

15 June CH, Warshauer JT, Bluestone JA. Is autoimmunity the Achilles' heel of cancer

immunotherapy? *Nat Med* 2017;**23**:540–7. doi:10.1038/nm.4321

16 Vesely MD, Kershaw MH, Schreiber RD, *et al.* Natural Innate and Adaptive Immunity to Cancer. *Annu Rev Immunol* 2011;**29**:235–71. doi:10.1146/annurev-immunol-031210-101324

17 Martins F, Sofiya L, Sykiotis GP, *et al.* Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. *Nat Rev Clin Oncol* 2019;**16**:563–80. doi:10.1038/s41571-019-0218-0

18 Carbone DP, Gandara DR, Antonia SJ, *et al.* Non-Small-Cell Lung Cancer: Role of the Immune System and Potential for Immunotherapy. *J Thorac Oncol* 2015;**10**:974–84. doi:10.1097/JTO.0000000000000551

19 Wong W, Yim YM, Kim A, *et al.* Assessment of costs associated with adverse events in patients with cancer. *PLoS One* 2018;**13**:e0196007. doi:10.1371/journal.pone.0196007

20 Diagnosis D of CT and. Common Terminology Criteria for Adverse Events. Natl. Cancer Inst. 2020.https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm (accessed 10 Jun 2020).

21 Jurczyk M, Król M, Midro A, *et al.* Cardiotoxicity of Fluoropyrimidines: Epidemiology, Mechanisms, Diagnosis, and Management. J. Clin. Med. . 2021;**10**. doi:10.3390/jcm10194426

22 Chong JH, Ghosh AK. Coronary Artery Vasospasm Induced by 5-fluorouracil: Proposed Mechanisms, Existing Management Options and Future Directions. *Interv Cardiol (London, England)* 2019;**14**:89–94. doi:10.15420/icr.2019.12

23 Jensen SA, Sørensen JB. Risk factors and prevention of cardiotoxicity induced by 5-fluorouracil or capecitabine. *Cancer Chemother Pharmacol* 2006;**58**:487–93. doi:10.1007/s00280-005-0178-1

24 Shahrokni A, Rajebi MR, Harold L SM. Cardiotoxicity of 5-Fluorouracil and Capecitabine in a Pancreatic Cancer Patient with a Novel Mutation in the Dihydropyrimidine Dehydrogenase Gene. *J Pancreas* 2009;**10**:215–20.

25 Lamberti M, Porto S, Marra M, *et al.* 5-Fluorouracil induces apoptosis in rat cardiocytes through intracellular oxidative stress. *J Exp Clin Cancer Res* 2012;**31**:60. doi:10.1186/1756-9966-31-60

26 Focaccetti C, Bruno A, Magnani E, *et al.* Effects of 5-fluorouracil on morphology, cell cycle, proliferation, apoptosis, autophagy and ROS production in endothelial cells and cardiomyocytes. *PLoS One* 2015;**10**:e0115686–e0115686. doi:10.1371/journal.pone.0115686

27 Chen DS, Mellman I. Oncology Meets Immunology: The Cancer-Immunity Cycle. *Immunity* 2013;**39**:1–10. doi:10.1016/j.immuni.2013.07.012

28 Khoja L, Day D, Wei-Wu Chen T, *et al.* Tumour- and class-specific patterns of immune-related adverse events of immune checkpoint inhibitors: a systematic review. *Ann Oncol* 2017;**28**:2377–85. doi:10.1093/annonc/mdx286

29 Haanen JBAG, Carbonnel F, Robert C, *et al.* Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol* 2017;**28**:iv119–42. doi:10.1093/annonc/mdx225

30 Electronic Medicines Compendium. Electronic Medicines Compendium (emc).

https://www.medicines.org.uk/emc/ (accessed 1 Apr 2021).

31 Larkin J, Chiarion-Sileni V, Gonzalez R, *et al.* Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med* 2015;**373**:23–34. doi:10.1056/NEJMoa1504030

32 Gandhi L, Rodríguez-Abreu D, Gadgeel S, *et al.* Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *N Engl J Med* 2018;**378**:2078–92. doi:10.1056/NEJMoa1801005

33 Wang Y, Zhou S, Yang F, *et al.* Treatment-Related Adverse Events of PD-1 and PD-L1 Inhibitors in Clinical Trials: A Systematic Review and Meta-analysis. *JAMA Oncol* 2019;**5**:1008–19. doi:10.1001/jamaoncol.2019.0393

34 Solinas C, Aiello M, Rozali E, *et al.* Programmed cell death-ligand 2: A neglected but important target in the immune response to cancer? *Transl Oncol* 2020;**13**:100811. doi:10.1016/j.tranon.2020.100811

35 Iwama S, De Remigis A, Callahan MK, *et al.* Pituitary Expression of CTLA-4 Mediates Hypophysitis Secondary to Administration of CTLA-4 Blocking Antibody. *Sci Transl Med* 2014;**6**:230ra45 LP-230ra45. doi:10.1126/scitranslmed.3008002

36 Wang DY, Salem J-E, Cohen J V, *et al.* Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. *JAMA Oncol* 2018;**4**:1721–8. doi:10.1001/jamaoncol.2018.3923

37 Chennamadhavuni A, Abushahin L, Jin N, *et al.* Risk Factors and Biomarkers for Immune-Related Adverse Events: A Practical Guide to Identifying High-Risk Patients and Rechallenging Immune Checkpoint Inhibitors. *Front Immunol* 2022;**13**.

doi:10.3389/fimmu.2022.779691

38 Samani A, Zhang S, Spiers L, *et al.* Impact of age on the toxicity of immune checkpoint inhibition. *J Immunother Cancer* 2020;**8**:e000871. doi:10.1136/jitc-2020-000871

39 Asada M, Mikami T, Niimura T, *et al.* The Risk Factors Associated with Immune Checkpoint Inhibitor-Related Pneumonitis. *Oncology* 2021;**99**:256–9. doi:10.1159/000512633

40 Triggianese P, Novelli L, Galdiero MR, *et al.* Immune checkpoint inhibitors-induced autoimmunity: The impact of gender. *Autoimmun Rev* 2020;**19**:102590. doi:https://doi.org/10.1016/j.autrev.2020.102590

41 Guzman-Prado Y, Ben Shimol J, Samson O. Body mass index and immune-related adverse events in patients on immune checkpoint inhibitor therapies: a systematic review and meta-analysis. *Cancer Immunol Immunother* 2021;**70**:89–100. doi:10.1007/s00262-020-02663-z

42 Okada N, Matsuoka R, Sakurada T, *et al.* Risk factors of immune checkpoint inhibitor-related interstitial lung disease in patients with lung cancer: a single-institution retrospective study. *Sci Rep* 2020;**10**:13773. doi:10.1038/s41598-020-70743-2

43 Atchley WT, Alvarez C, Saxena-Beem S, *et al.* Immune Checkpoint Inhibitor-Related Pneumonitis in Lung Cancer: Real-World Incidence, Risk Factors, and Management Practices Across Six Health Care Centers in North Carolina. *Chest* 2021;**160**:731–42. doi:10.1016/j.chest.2021.02.032

44 Tang S-Q, Tang L-L, Mao Y-P, *et al.* The Pattern of Time to Onset and Resolution of Immune-Related Adverse Events Caused by Immune Checkpoint Inhibitors in Cancer:

A Pooled Analysis of 23 Clinical Trials and 8,436 Patients. *Cancer Res Treat* 2021;**53**:339–54. doi:10.4143/crt.2020.790

45 Marthey L, Mateus C, Mussini C, *et al.* Cancer Immunotherapy with Anti-CTLA-4 Monoclonal Antibodies Induces an Inflammatory Bowel Disease. *J Crohns Colitis* 2016;**10**:395–401. doi:10.1093/ecco-jcc/jjv227

46 Calabrese LH, Calabrese C, Cappelli LC. Rheumatic immune-related adverse events from cancer immunotherapy. *Nat Rev Rheumatol* 2018;**14**:569–79. doi:10.1038/s41584-018-0074-9

47 Clotman K, Janssens K, Specenier P, *et al.* Programmed Cell Death-1 Inhibitor–Induced Type 1 Diabetes Mellitus. *J Clin Endocrinol Metab* 2018;**103**:3144–54. doi:10.1210/jc.2018-00728

48 Ericson-Neilsen W, Kaye AD. Steroids: pharmacology, complications, and practice delivery issues. *Ochsner J* 2014;**14**:203–7.

49 Schärfe CPI, Tremmel R, Schwab M, *et al.* Genetic variation in human drug-related genes. *Genome Med* 2017;**9**:117. doi:10.1186/s13073-017-0502-5

50 Hertz DL, Rae J. Pharmacogenetics of Cancer Drugs. *Annu Rev Med* 2015;**66**:65–81. doi:10.1146/annurev-med-053013-053944

51 Miteva-Marcheva NN, Ivanov HY, Dimitrov DK, *et al.* Application of pharmacogenetics in oncology. *Biomark Res* 2020;**8**:32. doi:10.1186/s40364-020-00213-4

52 Kirchhoff T, Ferguson R. Germline Genetics in Immuno-oncology: From Genome-Wide to Targeted Biomarker Strategies BT - Biomarkers for Immunotherapy of Cancer: Methods and Protocols. In: Thurin M, Cesano A, Marincola FM, eds. . New York, NY: :

Springer New York 2020. 93–117. doi:10.1007/978-1-4939-9773-2_4

53 Hertz DL, Rae J. Pharmacogenetics of Cancer Drugs. *Annu Rev Med* 2015;**66**:65–81. doi:10.1146/annurev-med-053013-053944

54 Roederer M, Quaye L, Mangino M, *et al.* The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. *Cell* 2015;**161**:387–403. doi:10.1016/j.cell.2015.02.046

55 Mangino M, Roederer M, Beddall MH, *et al.* Innate and adaptive immune traits are differentially affected by genetic and environmental factors. *Nat Commun* 2017;**8**:13850. doi:10.1038/ncomms13850

56 Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA CF and PH. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;:D1005–12.https://www.ebi.ac.uk/gwas/docs (accessed 30 Apr 2021).

57 Ueda H, Howson JMM, Esposito L, *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003;**423**:506–11. doi:10.1038/nature01621

58 Zeissig S, Petersen B-S, Tomczak M, *et al.* Early-onset Crohn's disease and autoimmunity associated with a variant in CTLA-4. *Gut* 2015;**64**:1889–97. doi:10.1136/gutjnl-2014-308541

59 Vogelsang M, Martinez CN, Rendleman J, *et al.* The Expression Quantitative Trait Loci

in Immune Pathways and their Effect on Cutaneous Melanoma Prognosis. *Clin Cancer Res* 2016;**22**:3268–80. doi:10.1158/1078-0432.CCR-15-2066

60 Gutierrez-Arcelus M, Rich SS, Raychaudhuri S. Autoimmune diseases - connecting risk alleles with molecular traits of the immune system. *Nat Rev Genet* 2016;**17**:160–74. doi:10.1038/nrg.2015.33

61 Parkes M, Cortes A, van Heel DA, *et al.* Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet* 2013;**14**:661–73. doi:10.1038/nrg3502

62 Seldin MF. The genetics of human autoimmune disease: A perspective on progress in the field and future directions. *J Autoimmun* 2015;**64**:1–12. doi:https://doi.org/10.1016/j.jaut.2015.08.015

63 Raychaudhuri S, Sandor C, Stahl EA, *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012;**44**:291–6. doi:10.1038/ng.1076

64 Noble JA, Erlich HA. Genetics of type 1 diabetes. *Cold Spring Harb Perspect Med* 2012;**2**:a007732–a007732. doi:10.1101/cshperspect.a007732

65 Martin MA, Kroetz DL. Abacavir pharmacogenetics--from initial reports to standard of care. *Pharmacotherapy* 2013;**33**:765–75. doi:10.1002/phar.1278

66 Pavlos R, Mallal S, Phillips E. HLA and pharmacogenetics of drug hypersensitivity. *Pharmacogenomics* 2012;**13**:1285–306. doi:10.2217/pgs.12.108

67 Zhernakova A, Eerligh P, Wijmenga C, *et al.* Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. *Genes Immun*

2005;**6**:459–61. doi:10.1038/sj.gene.6364220

68 Diaz-Gallo L-M, Espino-Paisán L, Fransen K, *et al.* Differential association of two PTPN22 coding variants with Crohn's disease and ulcerative colitis. *Inflamm Bowel Dis* 2011;**17**:2287–94. doi:10.1002/ibd.21630

69 Chung W-H, Hung S-I, Hong H-S, *et al.* A marker for Stevens–Johnson syndrome. *Nature* 2004;**428**:486. doi:10.1038/428486a

70 Lonjou C, Thomas L, Borot N, *et al.* A marker for Stevens-Johnson syndrome …: ethnicity matters. *Pharmacogenomics J* 2006;**6**:265–8. doi:10.1038/sj.tpj.6500356

71 Amstutz U, Henricks LM, Offer SM, *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clin Pharmacol Ther* 2018;**103**:210–6. doi:10.1002/cpt.911

72 White C, Scott RJ, Paul C, *et al.* Ethnic Diversity of DPD Activity and the DPYD Gene: Review of the Literature. *Pharmgenomics Pers Med* 2021;**14**:1603–17. doi:10.2147/PGPM.S337147

73 Johnson DB, Sullivan RJ, Menzies AM. Immune checkpoint inhibitors in challenging populations. *Cancer* 2017;**123**:1904–11. doi:10.1002/cncr.30642

74 Waterhouse P, Penninger JM, Timms E, *et al.* Lymphoproliferative Disorders with Early Lethality in Mice Deficient in Ctla-4. *Science (80-)* 1995;**270**:985–8. doi:10.1126/science.270.5238.985

75 Michot JM, Bigenwald C, Champiat S, *et al.* Immune-related adverse events with immune checkpoint blockade: a comprehensive review. *Eur J Cancer* 2016;**54**:139–48.

doi:https://doi.org/10.1016/j.ejca.2015.11.016

76 Khan SA, Pruitt SL, Xuan L, *et al.* Prevalence of Autoimmune Disease Among Patients With Lung Cancer: Implications for Immunotherapy Treatment Options. *JAMA Oncol* 2016;**2**:1507–8. doi:10.1001/jamaoncol.2016.2238

77 Menzies AM, Johnson DB, Ramanujam S, *et al.* Anti-PD-1 therapy in patients with advanced melanoma and preexisting autoimmune disorders or major toxicity with ipilimumab. *Ann Oncol* 2016;**28**:368–76. doi:10.1093/annonc/mdw443

78 Johnson DB, Sullivan RJ, Ott PA, *et al.* Ipilimumab Therapy in Patients With Advanced Melanoma and Preexisting Autoimmune Disorders. *JAMA Oncol* 2016;**2**:234–40. doi:10.1001/jamaoncol.2015.4368

79 What are genome-wide association studies (GWAS)? EMBL-EBI Train Online. 2020.https://www.ebi.ac.uk/training-beta/online/courses/gwas-catalogue-exploring-snp-trait-associations/what-is-gwas-catalog/what-are-genome-wide-association-studies-gwas/ (accessed 23 Jul 2020).

80 Tam V, Patel N, Turcotte M, *et al.* Benefits and limitations of genome-wide association studies. *Nat Rev Genet* 2019;**20**:467–84. doi:10.1038/s41576-019-0127-1

81 Li JH, Mazur CA, Berisa T, *et al.* Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. *Genome Res* 2021;**31**:529–37. doi:10.1101/gr.266486.120

82 Groha S, Alaiwi SA, Xu W, *et al.* Germline variants associated with toxicity to immune checkpoint blockade. *Nat Med* Published Online First: 2022. doi:10.1038/s41591-022-02094-6

83 Chat V, Ferguson R, Kirchhoff T. Germline genetic host factors as predictive biomarkers in immuno-oncology. *Immuno-Oncology Technol* 2019;**2**:14–21. doi:https://doi.org/10.1016/j.iotech.2019.08.001

84 Watanabe K, Taskesen E, van Bochoven A, *et al.* Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;**8**:1826. doi:10.1038/s41467-017-01261-5

85 Maranville JC, Cox NJ. Pharmacogenomic variants have larger effect sizes than genetic variants associated with other dichotomous complex traits. *Pharmacogenomics J* 2016;**16**:388–92. doi:10.1038/tpj.2015.47

86 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 2018;**19**:581–90. doi:10.1038/s41576-018-0018-x

87 Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 2020;**15**:2759–72. doi:10.1038/s41596-020-0353-1

88 Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* 2016;**17**:392–406. doi:10.1038/nrg.2016.27

89 Gmeiner WH. A narrative review of genetic factors affecting fluoropyrimidine toxicity. *Precis cancer Med* 2021;**4**:38. doi:10.21037/pcm-21-17

90 Mattison LK, Fourie J, Desmond RA, *et al.* Increased Prevalence of Dihydropyrimidine Dehydrogenase Deficiency in African-Americans Compared with Caucasians. *Clin Cancer Res* 2006;**12**:5491 LP – 5495. doi:10.1158/1078-0432.CCR-06-0747

91 van Kuilenburg ABP, Haasjes J, Richel DJ, *et al.* Clinical Implications of

Dihydropyrimidine Dehydrogenase (DPD) Deficiency in Patients with Severe 5-Fluorouracil-associated Toxicity: Identification of New Mutations in the DPD Gene. *Clin Cancer Res* 2000;**6**:4705 LP – 4712.http://clincancerres.aacrjournals.org/content/6/12/4705.abstract

92 Nie Q, Shrestha S, Tapper EE, *et al.* Quantitative Contribution of rs75017182 to Dihydropyrimidine Dehydrogenase mRNA Splicing and Enzyme Activity. *Clin Pharmacol Ther* 2017;**102**:662–70. doi:10.1002/cpt.685

93 Medicines Compendium E. 5-Fluorouracil injection, 50mg/ml, solution for injection. EMC. 2019.https://www.medicines.org.uk/emc/product/11013/smpc (accessed 23 Apr 2020).

94 Meulendijks D, Henricks LM, Sonke GS, *et al.* Clinical relevance of DPYD variants c.1679T>G, c.1236G>A/HapB3, and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *Lancet Oncol* 2015;**16**:1639–50. doi:10.1016/S1470-2045(15)00286-7

95 Amstutz U, Froehlich TK, Largiadèr CR. Dihydropyrimidine dehydrogenase gene as a major predictor of severe 5-fluorouracil toxicity. *Pharmacogenomics* 2011;**12**:1321–36. doi:10.2217/pgs.11.72

96 Deenen MJ, Meulendijks D, Boot H, *et al.* Phase 1a/1b and pharmacogenetic study of docetaxel, oxaliplatin and capecitabine in patients with advanced cancer of the stomach or the gastroesophageal junction. *Cancer Chemother Pharmacol* 2015;**76**:1285–95. doi:10.1007/s00280-015-2872-y

97 García-González X, Cortejoso L, García MI, *et al.* Variants in CDA and ABCB1 are

predictors of capecitabine-related adverse reactions in colorectal cancer. *Oncotarget; Vol 6, No 8* Published Online First: 2015.https://www.oncotarget.com/article/3289/text/

98 Joerger M, Huitema ADR, Boot H, *et al.* Germline TYMS genotype is highly predictive in patients with metastatic gastrointestinal malignancies receiving capecitabine-based chemotherapy. *Cancer Chemother Pharmacol* 2015;**75**:763–72. doi:10.1007/s00280-015-2698-7

99 Deenen MJ, Tol J, Burylo AM, *et al.* Relationship between Single Nucleotide Polymorphisms and Haplotypes in DPYD and Toxicity and Efficacy of Capecitabine in Advanced Colorectal Cancer. *Clin Cancer Res* 2011;**17**:3455–68. doi:10.1158/1078-0432.CCR-10-2209

100 O'Donnell PH, Trubetskoy V, Nurhussein-Patterson A, *et al.* Clinical evaluation of germline polymorphisms associated with capecitabine toxicity in breast cancer: TBCRC-015. *Breast Cancer Res Treat* 2020;**181**:623–33. doi:10.1007/s10549-020-05603-8

101 Rosmarin D, Palles C, Church D, *et al.* Genetic markers of toxicity from capecitabine and other fluorouracil-based regimens: investigation in the QUASAR2 study, systematic review, and meta-analysis. *J Clin Oncol* 2014;**32**:1031–9. doi:10.1200/JCO.2013.51.1857

102 Sharma R, Hoskins JM, Rivory LP, *et al.* Thymidylate Synthase and Methylenetetrahydrofolate Reductase Gene Polymorphisms and Toxicity to Capecitabine in Advanced Colorectal Cancer Patients. *Clin Cancer Res* 2008;**14**:817–25. doi:10.1158/1078-0432.CCR-07-0425

103 Largillier R, Etienne-Grimaldi M-C, Formento J-L, *et al.* Pharmacogenetics of Capecitabine in Advanced Breast Cancer Patients. *Clin Cancer Res* 2006;**12**:5496–502. doi:10.1158/1078-0432.CCR-06-0320

104 Zarate R, Rodríguez J, Bandres E, *et al.* Oxaliplatin, irinotecan and capecitabine as first-line therapy in metastatic colorectal cancer (mCRC): a dose-finding study and pharmacogenomic analysis. *Br J Cancer* 2010;**102**:987–94. doi:10.1038/sj.bjc.6605595

105 Meulendijks D, Jacobs BAW, Aliev A, *et al.* Increased risk of severe fluoropyrimidine-associated toxicity in patients carrying a G to C substitution in the first 28-bp tandem repeat of the thymidylate synthase 2R allele. *Int J Cancer* 2016;**138**:245–53. doi:https://doi.org/10.1002/ijc.29694

106 Lin S, Yue J, Guan X, *et al.* Polymorphisms of MTHFR and TYMS predict capecitabine-induced hand-foot syndrome in patients with metastatic breast cancer. *Cancer Commun (London, England)* 2019;**39**:57. doi:10.1186/s40880-019-0399-z

107 Loganayagam A, Arenas Hernandez M, Corrigan A, *et al.* Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *Br J Cancer* 2013;**108**:2505–15. doi:10.1038/bjc.2013.262

108 Caronia D, Martin M, Sastre J, *et al.* A Polymorphism in the Cytidine Deaminase Promoter Predicts Severe Capecitabine-Induced Hand-Foot Syndrome. *Clin Cancer Res* 2011;**17**:2006–13. doi:10.1158/1078-0432.CCR-10-1741

109 Martín M, Martínez N, Ramos M, *et al.* Standard Versus Continuous Administration of Capecitabine in Metastatic Breast Cancer (GEICAM/2009-05): A Randomized, Noninferiority Phase II Trial With a Pharmacogenetic Analysis. *Oncologist* 2015;**20**:111–2. doi:10.1634/theoncologist.2014-0379

110 Offer SM, Butterfield GL, Jerde CR, *et al.* microRNAs miR-27a and miR-27b Directly Regulate Liver Dihydropyrimidine Dehydrogenase Expression through Two Conserved Binding Sites. *Mol Cancer Ther* 2014;**13**:742 LP – 751. doi:10.1158/1535-7163.MCT-13-0878

111 Amstutz U, Offer SM, Sistonen J, *et al.* Polymorphisms in MIR27A Associated with Early-Onset Toxicity in Fluoropyrimidine-Based Chemotherapy. *Clin Cancer Res* 2015;**21**:2038 LP – 2044. doi:10.1158/1078-0432.CCR-14-2817

112 Wang Z, Sun X, Wang Y, *et al.* Association between miR-27a genetic variants and susceptibility to colorectal cancer. *Diagn Pathol* 2014;**9**:146. doi:10.1186/1746-1596-9-146

113 Zhu H, Wu H, Liu X, *et al.* Role of MicroRNA miR-27a and miR-451 in the regulation of MDR1/P-glycoprotein expression in human cancer cells. *Biochem Pharmacol* 2008;**76**:582–8. doi:10.1016/j.bcp.2008.06.007

114 Zeng Y, Yi R, Cullen BR. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J* 2005;**24**:138–48. doi:10.1038/sj.emboj.7600491

115 Meulendijks D, Henricks LM, Amstutz U, *et al.* Rs895819 in MIR27A improves the predictive value of DPYD variants to identify patients at risk of severe fluoropyrimidine-associated toxicity. *Int J Cancer* 2016;**138**:2752–61. doi:10.1002/ijc.30014

116 Yang X, Li G, Guan M, *et al.* Potential Gene Association Studies of Chemotherapy-Induced Cardiotoxicity: A Systematic Review and Meta-Analysis. *Front Cardiovasc Med* 2021;**8**. doi:10.3389/fcvm.2021.651269

117 Abdel-Wahab N, Diab A, Yu RK, *et al.* Genetic determinants of immune-related adverse events in patients with melanoma receiving immune checkpoint inhibitors. *Cancer Immunol Immunother* 2021;**70**:1939–49. doi:10.1007/s00262-020-02797-0

118 Bins S, Basak EA, El Bouazzaoui S, *et al.* Association between single-nucleotide polymorphisms and adverse events in nivolumab-treated non-small cell lung cancer patients. *Br J Cancer* 2018;**118**:1296–301. doi:10.1038/s41416-018-0074-1

119 Queirolo P, Dozin B, Morabito A, *et al.* CTLA-4 gene variant -1661A>G may predict the onset of endocrine adverse events in metastatic melanoma patients treated with ipilimumab. *Eur J Cancer* 2018;**97**:59–61. doi:https://doi.org/10.1016/j.ejca.2018.04.005

120 Cappelli LC, Dorak MT, Bettinotti MP, *et al.* Association of HLA-DRB1 shared epitope alleles and immune checkpoint inhibitor-induced inflammatory arthritis. *Rheumatology (Oxford)* 2019;**58**:476–80. doi:10.1093/rheumatology/key358

121 Magis Q, Gaudy-Marqueste C, Basire A, *et al.* Diabetes and Blood Glucose Disorders Under Anti-PD1. *J Immunother* 2018;**41**.https://journals.lww.com/immunotherapy-journal/Fulltext/2018/06000/Diabetes_and_Blood_Glucose_Disorders_Under.3.aspx

122 Refae S, Gal J, Ebran N, *et al.* Germinal Immunogenetics predict treatment outcome for PD-1/PD-L1 checkpoint inhibitors. *Invest New Drugs* Published Online First: 2019. doi:10.1007/s10637-019-00845-w

123 Iafolla MAJ, Yang C, Chandran V, *et al.* Predicting Toxicity and Response to Pembrolizumab Through Germline Genomic HLA Class 1 Analysis. *JNCI cancer Spectr* 2020;**5**:pkaa115–pkaa115. doi:10.1093/jncics/pkaa115

124 Hasan Ali O, Berner F, Bomze D, *et al.* Human leukocyte antigen variation is associated with adverse events of checkpoint inhibitors. *Eur J Cancer* 2019;**107**:8–14. doi:https://doi.org/10.1016/j.ejca.2018.11.009

125 Luo J, Martucci VL, Quandt Z, *et al.* Immunotherapy-Mediated Thyroid Dysfunction: Genetic Risk and Impact on Outcomes with PD-1 Blockade in Non–Small Cell Lung Cancer. *Clin Cancer Res* Published Online First: 7 July 2021. doi:10.1158/1078-0432.CCR-21-0921

126 Kirchhoff T, Ferguson R, Simpson D, *et al.* Germline determinants of immune related adverse events (irAEs) in melanoma immunotherapy response. *Ann Oncol* 2017;**28**:v407–8. doi:10.1093/annonc/mdx376.012

127 Montaudié H, Beranger GE, Reinier F, *et al.* Germline variants in exonic regions have limited impact on immune checkpoint blockade clinical outcomes in advanced melanoma. *Pigment Cell Melanoma Res* 2021;**34**:978–83. doi:https://doi.org/10.1111/pcmr.12958

128 Yano S, Ashida K, Sakamoto R, *et al.* Human leucocyte antigen DR15, a possible predictive marker for immune checkpoint inhibitor–induced secondary adrenal insufficiency. *Eur J Cancer* 2020;**130**:198–203. doi:https://doi.org/10.1016/j.ejca.2020.02.049

129 Marschner D, Falk M, Javorniczky NR, *et al.* MicroRNA-146a regulates immune-related adverse events caused by immune checkpoint inhibitors. *JCI insight* 2020;**5**:e132334. doi:10.1172/jci.insight.132334

130 Weidhaas J, Marco N, Scheffler AW, *et al.* Germline biomarkers predict toxicity to anti-PD1/PDL1 checkpoint therapy. *J Immunother Cancer* 2022;**10**:e003625.

doi:10.1136/jitc-2021-003625

131 Udagawa C, Nakano MH, Yoshida T, *et al.* Association between genetic variants and the risk of nivolumab-induced immune-related adverse events. *Pharmacogenomics* Published Online First: 21 October 2022. doi:10.2217/pgs-2022-0113

132 Goyette P, Boucher G, Mallon D, *et al.* High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet* 2015;**47**:172–9. doi:10.1038/ng.3176

133 Lowe JR, Perry DJ, Salama AKS, *et al.* Genetic risk analysis of a patient with fulminant autoimmune type 1 diabetes mellitus secondary to combination ipilimumab and nivolumab immunotherapy. *J Immunother cancer* 2016;**4**:89. doi:10.1186/s40425-016-0196-z

134 Groha S, Alaiwi SA, Xu W, *et al.* Germline variants associated with immunotherapy-related adverse events. *Prepr https//www.medrxiv.org/content/101101/2022041022273627v2* 2022;:2022.04.10.22273627.

135 Li H, Li Q, Sun S, *et al.* Integrated Bioinformatics Analysis Identifies ELAVL1 and APP as Candidate Crucial Genes for Crohn's Disease. *J Immunol Res* 2020;**2020**:3067273. doi:10.1155/2020/3067273

136 Abola M V, Prasad V, Jena AB. Association between treatment toxicity and outcomes in oncology clinical trials. *Ann Oncol Off J Eur Soc Med Oncol* 2014;**25**:2284–9. doi:10.1093/annonc/mdu444

137 Petrelli F, Grizzi G, Ghidini M, *et al.* Immune-related Adverse Events and Survival in

Solid Tumors Treated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-Analysis. *J Immunother* 2019;**43**:1. doi:10.1097/CJI.0000000000000300

138 Ricciuti B, Genova C, De Giglio A, *et al.* Impact of immune-related adverse events on survival in patients with advanced non-small cell lung cancer treated with nivolumab: long-term outcomes from a multi-institutional analysis. *J Cancer Res Clin Oncol* 2019;**145**:479–85. doi:10.1007/s00432-018-2805-3

139 Hua C, Boussemart L, Mateus C, *et al.* Association of Vitiligo With Tumor Response in Patients With Metastatic Melanoma Treated With Pembrolizumab. *JAMA Dermatology* 2016;**152**:45–51. doi:10.1001/jamadermatol.2015.2707

140 Salgado J, Zabalegui N, Gil C, *et al.* Polymorphisms in the thymidylate synthase and dihydropyrimidine dehydrogenase genes predict response and toxicity to capecitabine-raltitrexed in colorectal cancer. *Oncol Rep* 2007;**17**:325–8. doi:10.3892/or.17.2.325

141 Gao J, He Q, Hua D, *et al.* Polymorphism of TS 3'-UTR predicts survival of Chinese advanced gastric cancer patients receiving first-line capecitabine plus paclitaxel. *Clin Transl Oncol* 2013;**15**:619–25. doi:10.1007/s12094-012-0979-8

142 Chat V, Ferguson R, Simpson D, *et al.* Autoimmune genetic risk variants as germline biomarkers of response to melanoma immune-checkpoint inhibition. *Cancer Immunol Immunother* 2019;**68**:897–905. doi:10.1007/s00262-019-02318-8

143 Simoncic PD, Lee-Loy A, Barber DL, *et al.* The T Cell Protein Tyrosine Phosphatase Is a Negative Regulator of Janus Family Kinases 1 and 3. *Curr Biol* 2002;**12**:446–53. doi:10.1016/S0960-9822(02)00697-8

144 Gutierrez-Arcelus M, Rich SS, Raychaudhuri S. Autoimmune diseases - connecting risk alleles with molecular traits of the immune system. *Nat Rev Genet* 2016;**17**:160–74. doi:10.1038/nrg.2015.33

145 Nomizo T, Ozasa H, Tsuji T, *et al.* Clinical Impact of Single Nucleotide Polymorphism in PD-L1 on Response to Nivolumab for Advanced Non-Small-Cell Lung Cancer Patients. *Sci Rep* 2017;**7**:45124. doi:10.1038/srep45124

146 Glimelius B, Garmo H, Berglund Å, *et al.* Prediction of irinotecan and 5-fluorouracil toxicity and response in patients with advanced colorectal cancer. *Pharmacogenomics J* 2011;**11**:61–71. doi:10.1038/tpj.2010.10

147 Henriksen TV, Tarazona N, Frydendahl A, *et al.* Circulating Tumor DNA in Stage III Colorectal Cancer, beyond Minimal Residual Disease Detection, toward Assessment of Adjuvant Therapy Efficacy and Clinical Behavior of Recurrences. *Clin Cancer Res* 2022;**28**:507–17. doi:10.1158/1078-0432.CCR-21-2404

148 Parikh AR, Mojtahed A, Schneider JL, *et al.* Serial ctDNA Monitoring to Predict Response to Systemic Therapy in Metastatic Gastrointestinal Cancers. *Clin Cancer Res* 2020;**26**:1877–85. doi:10.1158/1078-0432.CCR-19-3467

149 Meulendijks D, Henricks LM, Jacobs BAW, *et al.* Pretreatment serum uracil concentration as a predictor of severe and fatal fluoropyrimidine-associated toxicity. *Br J Cancer* 2017;**116**:1415–24. doi:10.1038/bjc.2017.94

150 Teng F, Meng X, Kong L, *et al.* Progress and challenges of predictive biomarkers of anti PD-1/PD-L1 immunotherapy: A systematic review. *Cancer Lett* 2018;**414**:166–73. doi:https://doi.org/10.1016/j.canlet.2017.11.014

151 Lim SY, Lee JH, Gide TN, *et al.* Circulating Cytokines Predict Immune-Related Toxicity in Melanoma Patients Receiving Anti-PD-1–Based Immunotherapy. *Clin Cancer Res* 2019;**25**:1557 LP – 1563. doi:10.1158/1078-0432.CCR-18-2795

152 Ozawa Y, Amano Y, Kanata K, *et al.* Impact of early inflammatory cytokine elevation after commencement of PD-1 inhibitors to predict efficacy in patients with non-small cell lung cancer. *Med Oncol* 2019;**36**:33. doi:10.1007/s12032-019-1255-3

153 Tarhini AA, Zahoor H, Lin Y, *et al.* Baseline circulating IL-17 predicts toxicity while TGF-β1 and IL-10 are prognostic of relapse in ipilimumab neoadjuvant therapy of melanoma. *J Immunother cancer* 2015;**3**:39. doi:10.1186/s40425-015-0081-1

154 Oh DY, Cham J, Zhang L, *et al.* Immune Toxicities Elicted by CTLA-4 Blockade in Cancer Patients Are Associated with Early Diversification of the T-cell Repertoire. *Cancer Res* 2017;**77**:1322–30. doi:10.1158/0008-5472.CAN-16-2324

155 Das R, Bar N, Ferreira M, *et al.* Early B cell changes predict autoimmunity following combination immune checkpoint blockade. *J Clin Invest* 2018;**128**:715–20. doi:10.1172/JCI96798

156 Patel AJ, Willsmore ZN, Khan N, *et al.* Regulatory B cell repertoire defects predispose lung cancer patients to immune-related toxicity following checkpoint blockade. *Nat Commun* 2022;**13**:3148. doi:10.1038/s41467-022-30863-x

157 Anderson R, Theron AJ, Rapoport BL. Immunopathogenesis of Immune Checkpoint Inhibitor-Related Adverse Events: Roles of the Intestinal Microbiome and Th17 Cells. *Front Immunol* 2019;**10**:2254. doi:10.3389/fimmu.2019.02254

158 Chaput N, Lepage P, Coutzac C, *et al.* Baseline gut microbiota predicts clinical response

and colitis in metastatic melanoma patients treated with ipilimumab. *Ann Oncol* 2017;**28**:1368–79. doi:10.1093/annonc/mdx108

159 Kerr RS, Love S, Segelov E, *et al.* Adjuvant capecitabine plus bevacizumab versus capecitabine alone in patients with colorectal cancer (QUASAR 2): an open-label, randomised phase 3 trial. *Lancet Oncol* 2016;**17**:1543–57. doi:10.1016/S1470-2045(16)30172-3

160 Palles C, Fotheringham S, Chegwidden L, *et al.* An Evaluation of the Diagnostic Accuracy of a Panel of Variants in DPYD and a Single Variant in ENOSF1 for Predicting Common Capecitabine Related Toxicities. *Cancers (Basel)* 2021;**13**. doi:10.3390/cancers13071497

161 Maughan TS, Adams RA, Smith CG, *et al.* Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet* 2011;**377**:2103–14. doi:10.1016/S0140-6736(11)60613-2

162 Iveson TJ, Kerr RS, Saunders MP, *et al.* 3 versus 6 months of adjuvant oxaliplatin-fluoropyrimidine combination therapy for colorectal cancer (SCOT): an international, randomised, phase 3, non-inferiority trial. *Lancet Oncol* 2018;**19**:562–78. doi:10.1016/S1470-2045(18)30093-7

163 Deenen MJ, Meulendijks D, Cats A, *et al.* Upfront Genotyping of DPYD*2A to Individualize Fluoropyrimidine Therapy: A Safety and Cost Analysis. *J Clin Oncol* 2015;**34**:227–34. doi:10.1200/JCO.2015.63.1325

164 Marchini J, Howie B, Myers S, *et al.* A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906–13.

doi:10.1038/ng2088

165 Freeman C MJ. GTOOL. 2012.https://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html (accessed 1 Apr 2023).

166 Sebastiani P, Timofeev N, Dworkis DA, *et al.* Genome-wide association studies and the genetic dissection of complex traits. *Am J Hematol* 2009;**84**:504–15. doi:https://doi.org/10.1002/ajh.21440

167 Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;**11**:499–511. doi:10.1038/nrg2796

168 Rizvi A, Karaesmen E, Morgan M S-CL. gwasurvivr: gwasurvivr: an R package for genome wide survival analysis. R package version 1.14.0. 2022. doi:10.18129/B9.bioc.gwasurvivr

169 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**:e164–e164. doi:10.1093/nar/gkq603

170 Earp MA, Goode EL. Linkage Disequilibrium BT - Encyclopedia of Cancer. In: Schwab M, ed. . Berlin, Heidelberg: : Springer Berlin Heidelberg 2017. 2496–502. doi:10.1007/978-3-662-46875-3_3368

171 Yang J, Ferreira T, Morris AP, *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012;**44**:369–75. doi:10.1038/ng.2213

172 Liberzon A, Subramanian A, Pinchback R, *et al.* Molecular signatures database

(MSigDB) 3.0. *Bioinformatics* 2011;**27**:1739–40. doi:10.1093/bioinformatics/btr260

173 Broad Institute of MIT and Harvard. GTEX PORTAL. 2021. https://gtexportal.org/home/ (accessed 3 Sep 2022).

174 Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc London Ser B, Biol Sci* 2013;**368**:20120362. doi:10.1098/rstb.2012.0362

175 Võsa U, Claringbould A, Westra H-J, *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* 2021;**53**:1300–10. doi:10.1038/s41588-021-00913-z

176 GSE87112 | GEO. A Compendium of Chromatin Contact Maps Reveal Spatially Active Regions in the Human Genome. 2016.https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112.

177 Giusti-Rodríguez P, Lu L, Yang Y, *et al.* Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits. *bioRxiv* 2019;:406330. doi:10.1101/406330

178 Bastarache L, Denny JC, Roden DM. Phenome-Wide Association Studies. *JAMA* 2022;**327**:75–6. doi:10.1001/jama.2021.20356

179 Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* 2019;**8**:giz082. doi:10.1093/gigascience/giz082

180 Robin X, Turck N, Hainard A, *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-2105-12-77

181 Fox J, Weisber S. *An R Companion to Applied Regression*. Third Edit. Thousand Oaks

CA: : SAGE Publications 2019.

https://socialsciences.mcmaster.ca/jfox/Books/Companion/

182 Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics* 2020;**36**:5424–31. doi:10.1093/bioinformatics/btaa1029

183 Wainberg M, Sinnott-Armstrong N, Mancuso N, *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* 2019;**51**:592–9. doi:10.1038/s41588-019-0385-z

184 Kent WJ, Sugnet CW, Furey TS, *et al.* The Human Genome Browser at UCSC. *Genome Res* 2002;**12**:996–1006. doi:10.1101/gr.229102

185 Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**:263–5. doi:10.1093/bioinformatics/bth457

186 Tawfik E, Ahamed M, Almalik A, *et al.* Prolonged exposure of colon cancer cells to 5-fluorouracil nanoparticles improves its anticancer activity. *Saudi Pharm J* 2017;**25**:206–13. doi:https://doi.org/10.1016/j.jsps.2016.05.010

187 Ariyoshi M, Katane M, Hamase K, *et al.* D-Glutamate is metabolized in the heart mitochondria. *Sci Rep* 2017;**7**:43911. doi:10.1038/srep43911

188 Nonnenmacher Y, Palorini R, Hiller K. Determining Compartment-Specific Metabolic Fluxes BT - Metabolic Signaling: Methods and Protocols. In: Fendt S-M, Lunt SY, eds. . New York, NY: : Springer New York 2019. 137–49. doi:10.1007/978-1-4939-8769-6_10

189 Bonacorsi S, Visseaux B, Bouzid D, *et al.* Systematic Review on the Correlation of Quantitative PCR Cycle Threshold Values of Gastrointestinal Pathogens With Patient Clinical Presentation and Outcomes. Front. Med. 2021;**8**:711809.

doi:10.3389/fmed.2021.711809

190 Delaneau O, Howie B, Cox AJ, *et al.* Haplotype Estimation Using Sequencing Reads. *Am J Hum Genet* 2013;**93**:687–96. doi:10.1016/j.ajhg.2013.09.002

191 Jia X, Han B, Onengut-Gumuscu S, *et al.* Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLoS One* 2013;**8**:e64683.https://doi.org/10.1371/journal.pone.0064683

192 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.https://arxiv.org/abs/1303.3997

193 Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;**581**:434–43. doi:10.1038/s41586-020-2308-7

194 Yen JL, McLeod HL. Should DPD analysis be required prior to prescribing fluoropyrimidines? *Eur J Cancer* 2007;**43**:1011–6. doi:https://doi.org/10.1016/j.ejca.2007.01.030

195 Branda RF, Nigels E, Lafayette AR, *et al.* Nutritional Folate Status Influences the Efficacy and Toxicity of Chemotherapy in Rats. *Blood* 1998;**92**:2471–6. doi:https://doi.org/10.1182/blood.V92.7.2471

196 Hodroj K, Barthelemy D, Lega J-C, *et al.* Issues and limitations of available biomarkers for fluoropyrimidine-based chemotherapy toxicity, a narrative review of the literature. *ESMO open* 2021;**6**:100125. doi:10.1016/j.esmoop.2021.100125

197 Boisdron-Celle M, Remaud G, Traore S, *et al.* 5-Fluorouracil-related severe toxicity: A comparison of different methods for the pretherapeutic detection of

dihydropyrimidine dehydrogenase deficiency. *Cancer Lett* 2007;**249**:271–82. doi:https://doi.org/10.1016/j.canlet.2006.09.006

198 Petrelli F, Cabiddu M, Barni S. 5-Fluorouracil or capecitabine in the treatment of advanced colorectal cancer: a pooled-analysis of randomized trials. *Med Oncol* 2012;**29**:1020–9. doi:10.1007/s12032-011-9958-0

199 Leonard R, Hennessy BT, Blum JL, *et al.* Dose-Adjusting Capecitabine Minimizes Adverse Effects While Maintaining Efficacy: A Retrospective Review of Capecitabine for Metastatic Breast Cancer. *Clin Breast Cancer* 2011;**11**:349–56. doi:10.1016/j.clbc.2011.06.005

200 Gieschke R, Burger H-U, Reigner B, *et al.* Population pharmacokinetics and concentration-effect relationships of capecitabine metabolites in colorectal cancer patients. *Br J Clin Pharmacol* 2003;**55**:252–63. doi:10.1046/j.1365-2125.2003.01765.x

201 Song B, Yan G, Hao H, *et al.* rs11671784 G/A and rs895819 A/G Polymorphisms Inversely Affect Gastric Cancer Susceptibility and miR-27a Expression in a Chinese Population. *Med Sci Monit* 2014;**20**:2318–26. doi:https://dx.doi.org/10.12659/MSM.892499

202 Yang Q, Jie Z, Ye S, *et al.* Genetic variations in miR-27a gene decrease mature miR-27a level and reduce gastric cancer susceptibility. *Oncogene* 2014;**33**:193–202. doi:10.1038/onc.2012.569

203 Froehlich TK, Amstutz U, Aebi S, *et al.* Clinical importance of risk variants in the dihydropyrimidine dehydrogenase gene for the prediction of early-onset fluoropyrimidine toxicity. *Int J Cancer* 2015;**136**:730–9. doi:10.1002/ijc.29025

204 Bae H, Perls T, Steinberg M, *et al.* Bayesian Polynomial Regression Models to Fit Multiple Genetic Models for Quantitative Traits. *Bayesian Anal* 2015;**10**:53–74. doi:10.1214/14-BA880

205 Horita N, Kaneko T. Genetic model selection for a case–control study and a meta-analysis. *Meta Gene* 2015;**5**:1–8. doi:https://doi.org/10.1016/j.mgene.2015.04.003

206 Catucci I, Osorio A, Arver B, *et al.* Individuals with FANCM biallelic mutations do not develop Fanconi anemia, but show risk for breast cancer, chemotherapy toxicity and may display chromosome fragility. *Genet Med* 2018;**20**:452–7. doi:10.1038/gim.2017.123

207 SLC22A5 solute carrier family 22 member 5 [Homo sapiens (human)]. Natl. Libr. Med. 2022.https://www.ncbi.nlm.nih.gov/gene/6584#summary (accessed 9 Jan 2023).

208 Fairfax BP, Makino S, Radhakrishnan J, *et al.* Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles. *Nat Genet* 2012;**44**:502–10. doi:10.1038/ng.2205

209 Fairfax BP, Humburg P, Makino S, *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science (80-)* 2014;**343**:1246949. doi:10.1126/science.1246949

210 GTEx Portal. Broad Inst. MIT Harvard. 2019.https://www.gtexportal.org/home/ (accessed 22 Jul 2020).

211 de Leeuw CA, Mooij JM, Heskes T, *et al.* MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015;**11**:e1004219. doi:10.1371/journal.pcbi.1004219

212 Lee HN, Jeong MS, Jang SB. Molecular Characteristics of Amyloid Precursor Protein

(APP) and Its Effects in Cancer. *Int J Mol Sci* 2021;**22**. doi:10.3390/ijms22094999

213 Chen M-H, Raffield LM, Mousas A, *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* 2020;**182**:1198-1213.e14. doi:https://doi.org/10.1016/j.cell.2020.06.045

214 Elsworth B, Lyon M, Alexander T, *et al.* The MRC IEU OpenGWAS data infrastructure. *bioRxiv* 2020;:2020.08.10.244293. doi:10.1101/2020.08.10.244293

215 Elahimanesh M, Najafi M. Cross talk between bacterial and human gene networks enriched using ncRNAs in IBD disease. *Sci Rep* 2023;**13**:7704. doi:10.1038/s41598-023-34780-x

216 Polk A, Vaage-Nilsen M, Vistisen K, *et al.* Cardiotoxicity in cancer patients treated with 5-fluorouracil or capecitabine: A systematic review of incidence, manifestations and predisposing factors. *Cancer Treat Rev* 2013;**39**:974–84. doi:https://doi.org/10.1016/j.ctrv.2013.03.005

217 Chang VY, Wang JJ. Pharmacogenetics of Chemotherapy-Induced Cardiotoxicity. *Curr Oncol Rep* 2018;**20**:52. doi:10.1007/s11912-018-0696-8

218 Dickel DE, Zhu Y, Nord AS, *et al.* Function-based identification of mammalian enhancers using site-specific integration. *Nat Methods* 2014;**11**:566–71. doi:10.1038/nmeth.2886

219 Montefiori LE, Sobreira DR, Sakabe NJ, *et al.* A promoter interaction map for cardiovascular disease genetics. *Elife* 2018;**7**:e35788. doi:10.7554/eLife.35788

220 Schlosser P, Li Y, Sekula P, *et al.* Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. *Nat Genet* 2020;**52**:167–76.

doi:10.1038/s41588-019-0567-8

221 Katane M, Ariyoshi M, Tateishi S, *et al.* Structural and enzymatic properties of mammalian d-glutamate cyclase. *Arch Biochem Biophys* 2018;**654**:10–8. doi:https://doi.org/10.1016/j.abb.2018.07.005

222 Tomita R, Todoroki K, Machida K, *et al.* Assessment of the Efficacy of Anticancer Drugs by Amino Acid Metabolomics Using Fluorescence Derivatization-HPLC. *Anal Sci* 2014;**30**:751–8. doi:10.2116/analsci.30.751

223 Goncharov N V, Jenkins RO, Radilov AS. Toxicology of fluoroacetate: a review, with possible directions for therapy research. *J Appl Toxicol* 2006;**26**:148–61. doi:https://doi.org/10.1002/jat.1118

224 Lischke J, Lang C, Sawodny O, *et al.* Impairment of energy metabolism in cardiomyocytes caused by 5-FU catabolites can be compensated by administration of amino acids. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2015. 5363–6. doi:10.1109/EMBC.2015.7319603

225 Claycomb WC, Lanson NA, Stallworth BS, *et al.* HL-1 cells: A cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc Natl Acad Sci* 1998;**95**:2979–84. doi:10.1073/pnas.95.6.2979

226 Saif MW, Smith M, Maloney A. The First Case of Severe Takotsubo Cardiomyopathy Associated with 5-Fluorouracil in a Patient with Abnormalities of Both Dihydropyrimidine Dehydrogenase (DPYD) and Thymidylate Synthase (TYMS) Genes. Cureus. 2016;**8**:e783. doi:10.7759/cureus.783

227 National Library of Medicine. TPK1 thiamin pyrophosphokinase 1 [Homo sapiens (human)]. 2022.

228 Polat B, Suleyman H, Sener E, *et al.* Examination of the Effects of Thiamine and Thiamine Pyrophosphate on Doxorubicin-Induced Experimental Cardiotoxicity. *J Cardiovasc Pharmacol Ther* 2014;**20**:221–9. doi:10.1177/1074248414552901

229 Kattoor AJ, Goel A, Mehta JL. Thiamine Therapy for Heart Failure: a Promise or Fiction? *Cardiovasc Drugs Ther* 2018;**32**:313–7. doi:10.1007/s10557-018-6808-8

230 Coskun R, Turan MI, Turan IS, *et al.* The protective effect of thiamine pyrophosphate, but not thiamine, against cardiotoxicity induced with cisplatin in rats. *Drug Chem Toxicol* 2014;**37**:290–4. doi:10.3109/01480545.2013.851688

231 Kurki MI, Karjalainen J, Palta P, *et al.* FinnGen: Unique genetic insights from combining isolated population and national health register data. *medRxiv* 2022;:2022.03.03.22271360. doi:10.1101/2022.03.03.22271360

232 Pan-UKB team. Pan-ancestry genetic analysis of the UK BioBank performed at the Broad Institute. 2020.https://pan.ukbb.broadinstitute.org/ (accessed 12 Mar 2023).

233 Churchhouse, C and Neale B. Rapis GWAS Of Thousands OF PHENOTYPES FOR 337,000 SAMPLES IN THE UK BIOBANKTitle. UKBiobank. 2017.http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank (accessed 13 Mar 2023).

234 Larkin J, Chiarion-Sileni V, Gonzalez R, *et al.* Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med* 2019;**381**:1535–46. doi:10.1056/NEJMoa1910836

235 Lee KA, Thomas AM, Bolte LA, *et al.* Cross-cohort gut microbiome associations with immune checkpoint inhibitor response in advanced melanoma. *Nat Med* 2022;**28**:535–44. doi:10.1038/s41591-022-01695-5

236 Boyle AP, Hong EL, Hariharan M, *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;**22**:1790–7. doi:10.1101/gr.137323.112

237 Chen L, Ge B, Casale FP, *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* 2016;**167**:1398-1414.e24. doi:10.1016/j.cell.2016.10.026

238 Schmitt AD, Hu M, Jung I, *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* 2016;**17**:2042–59. doi:10.1016/j.celrep.2016.10.061

239 Schwartzentruber J, Foskolou S, Kilpinen H, *et al.* Molecular and functional variation in iPSC-derived sensory neurons. *Nat Genet* 2018;**50**:54–61. doi:10.1038/s41588-017-0005-8

240 Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, *et al.* Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* 2013;**2**:e00523. doi:10.7554/eLife.00523

241 Lappalainen T, Sammeth M, Friedländer MR, *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;**501**:506–11. doi:10.1038/nature12531

242 Momozawa Y, Dmitrieva J, Théâtre E, *et al.* IBD risk loci are enriched in multigenic

regulatory modules encompassing putative causative genes. *Nat Commun* 2018;**9**:2427. doi:10.1038/s41467-018-04365-8

243 Jaffe AE, Straub RE, Shin JH, *et al.* Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci* 2018;**21**:1117–25. doi:10.1038/s41593-018-0197-y

244 van der Net JB, Janssens ACJW, Eijkemans MJC, *et al.* Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies. *Eur J Hum Genet* 2008;**16**:1111–6. doi:10.1038/ejhg.2008.59

245 UniProt. Q7Z3B1 · NEGR1_HUMAN. https://www.uniprot.org/uniprotkb/Q7Z3B1/entry#function (accessed 29 Oct 2023).

246 Verlouw JAM, Clemens E, de Vries JH, *et al.* A comparison of genotyping arrays. *Eur J Hum Genet* 2021;**29**:1611–24. doi:10.1038/s41431-021-00917-7

247 Du C, III JB, Cook R, *et al.* MICA immune complex formed with alpha 3 domain-specific antibody activates human NK cells in a Fc-dependent manner. *J Immunother Cancer* 2019;**7**:207. doi:10.1186/s40425-019-0687-9

248 Safran M, Rosen N, Twik M, BarShir R, Iny Stein T, Dahary D, Fishilevich S and LD. GeneCards – the human gene database. GeneCards Suite Chapter, Pract. Guid. to Life Sci. Databases. 2022;:27–56.www.genecards.org (accessed 18 Oct 2022).

249 Chin IS, Khan A, Olsson-Brown A, *et al.* Germline genetic variation and predicting immune checkpoint inhibitor induced toxicity. *npj Genomic Med* 2022;**7**:73. doi:10.1038/s41525-022-00345-6

250 Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores.

Hum Mol Genet 2019;**28**:R133–42. doi:10.1093/hmg/ddz187

251 Thomas M, Sakoda LC, Hoffmeister M, *et al.* Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am J Hum Genet* 2020;**107**:432–44. doi:10.1016/j.ajhg.2020.07.006

252 Orrù V, Steri M, Sidore C, *et al.* Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. *Nat Genet* 2020;**52**:1036–45. doi:10.1038/s41588-020-0684-4

253 Sun BB, Maranville JC, Peters JE, *et al.* Genomic atlas of the human plasma proteome. *Nature* 2018;**558**:73–9. doi:10.1038/s41586-018-0175-2

254 Kanai M, Akiyama M, Takahashi A, *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* 2018;**50**:390–400. doi:10.1038/s41588-018-0047-6

255 Medici M, Porcu E, Pistis G, *et al.* Identification of novel genetic Loci associated with thyroid peroxidase antibodies and clinical thyroid disease. *PLoS Genet* 2014;**10**:e1004123. doi:10.1371/journal.pgen.1004123

256 Motzer RJ, Tannir NM, McDermott DF, *et al.* Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *N Engl J Med* 2018;**378**:1277–90. doi:10.1056/NEJMoa1712126

257 Ahola-Olli A V, Würtz P, Havulinna AS, *et al.* Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am J Hum Genet* 2017;**100**:40–50. doi:10.1016/j.ajhg.2016.11.007

258 ZNF701. Hum. Protein Atlas. https://www.proteinatlas.org/ENSG00000167562-

ZNF701/immune+cell (accessed 31 Dec 2022).

259 Labbé C, Goyette P, Lefebvre C, *et al.* MAST3: a novel IBD risk factor that modulates TLR4 signaling. *Genes Immun* 2008;**9**:602–12. doi:10.1038/gene.2008.57

260 Zhu H, Huang S, Yue M, *et al.* Dysregulated Up-Frameshift Protein 1 Promotes Ulcerative Colitis Pathogenesis Through the TNFR1-NF-κB/MAPKs Pathway. *Dig Dis Sci* 2018;**63**:2593–603. doi:10.1007/s10620-018-5171-8

261 Ghosh C, Xing Y, Li S, *et al.* Sorting nexin 6 interacts with Cullin3 and regulates programmed death ligand 1 expression. *FEBS Lett* 2021;**595**:2558–69. doi:https://doi.org/10.1002/1873-3468.14191

262 Hong M, Ye BD, Yang S-K, *et al.* Immunochip Meta-Analysis of Inflammatory Bowel Disease Identifies Three Novel Loci and Four Novel Associations in Previously Reported Loci. *J Crohn's Colitis* 2018;**12**:730–41. doi:10.1093/ecco-jcc/jjy002

263 Sakurai T, Nishiyama H, Sakai K, *et al.* Mucosal microbiota and gene expression are associated with long-term remission after discontinuation of adalimumab in ulcerative colitis. *Sci Rep* 2020;**10**:19186. doi:10.1038/s41598-020-76175-2

264 Abuhammour W, Yavuz L, Jain R, *et al.* Genetic and Clinical Characteristics of Patients in the Middle East With Multisystem Inflammatory Syndrome in Children. *JAMA Netw Open* 2022;**5**:e2214985–e2214985. doi:10.1001/jamanetworkopen.2022.14985

265 Marukian S, Andrus L, Sheahan TP, *et al.* Hepatitis C virus induces interferon-λ and interferon-stimulated genes in primary liver cultures. *Hepatology* 2011;**54**:1913–23. doi:https://doi.org/10.1002/hep.24580

266 Pan H, Wang X, Huang W, *et al.* Interferon-Induced Protein 44 Correlated With

Immune Infiltration Serves as a Potential Prognostic Indicator in Head and Neck Squamous Cell Carcinoma. *Front Oncol* 2020;**10**. doi:10.3389/fonc.2020.557157

267 DeDiego ML, Nogales A, Martinez-Sobrido L, *et al.* Interferon-Induced Protein 44 Interacts with Cellular FK506-Binding Protein 5, Negatively Regulates Host Antiviral Responses, and Supports Virus Replication. *MBio* 2019;**10**. doi:10.1128/mBio.01839-19

268 Ovsyannikova AK, Rymar OD, Shakhtshneider E V, *et al.* ABCC8-Related Maturity-Onset Diabetes of the Young (MODY12): Clinical Features and Treatment Perspective. *Diabetes Ther Res Treat Educ diabetes Relat Disord* 2016;**7**:591–600. doi:10.1007/s13300-016-0192-9

269 Lin L, Quan H, Chen K, *et al.* ABCC8-Related Maturity-Onset Diabetes of the Young (MODY12): A Report of a Chinese Family . Front. Endocrinol. . 2020;**11**.https://www.frontiersin.org/articles/10.3389/fendo.2020.00645

270 Shima KR, Usuda R, Futatani T, *et al.* Heterogeneous nature of diabetes in a family with a gain-of-function mutation in the ATP-binding cassette subfamily C member 8 (ABCC8) gene. *Endocr J* 2018;**65**:1055–9. doi:10.1507/endocrj.EJ18-0054

271 Ge Y, Chen Z, Fu Y, *et al.* Identification and validation of hub genes of synovial tissue for patients with osteoarthritis and rheumatoid arthritis. *Hereditas* 2021;**158**:37. doi:10.1186/s41065-021-00201-0

272 Servitja J-M, Pignatelli M, Maestro MÁ, *et al.* Hnf1α (MODY3) Controls Tissue-Specific Transcriptional Programs and Exerts Opposed Effects on Cell Growth in Pancreatic Islets and Liver. *Mol Cell Biol* 2009;**29**:2945–59. doi:10.1128/MCB.01389-08

273 Zhang E, Huang X, He J. Integrated bioinformatic analysis of HNF1A in human cancers. *J Int Med Res* 2021;**49**:0300060521997326. doi:10.1177/0300060521997326

274 Trézéguet V, Fatrouni H, Merched AJ. Immuno-Metabolic Modulation of Liver Oncogenesis by the Tryptophan Metabolism. *Cells* 2021;**10**. doi:10.3390/cells10123469

275 Zhong L, Wu Q, Chen F, *et al.* Immune-related adverse events: promising predictors for efficacy of immune checkpoint inhibitors. *Cancer Immunol Immunother* 2021;**70**:2559–76. doi:10.1007/s00262-020-02803-5

276 Sun Q, Sun H, Wu N, *et al.* Patients with melanoma treated with immune checkpoint inhibitors who had non-thyroid endocrine and skin immune-related adverse events have better prognosis: A systematic review and meta-analysis . Front. Oncol. . 2022;**12**.https://www.frontiersin.org/articles/10.3389/fonc.2022.976224

277 Hussaini S, Chehade R, Boldt RG, *et al.* Association between immune-related side effects and efficacy and benefit of immune checkpoint inhibitors – A systematic review and meta-analysis. *Cancer Treat Rev* 2021;**92**:102134. doi:https://doi.org/10.1016/j.ctrv.2020.102134

278 Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases . Front. Genet. . 2020;**11**.https://www.frontiersin.org/articles/10.3389/fgene.2020.00424

279 Shrestha S, Zhang C, Jerde CR, *et al.* Gene-Specific Variant Classifier (DPYD-Varifier) to Identify Deleterious Alleles of Dihydropyrimidine Dehydrogenase. *Clin Pharmacol Ther* 2018;**104**:709–18. doi:https://doi.org/10.1002/cpt.1020

280 Teulings H-E, Limpens J, Jansen SN, *et al.* Vitiligo-Like Depigmentation in Patients With Stage III-IV Melanoma Receiving Immunotherapy and Its Association With Survival: A Systematic Review and Meta-Analysis. *J Clin Oncol* 2015;**33**:773–81. doi:10.1200/JCO.2014.57.4756

281 Uozumi S, Enokida T, Suzuki S, *et al.* Predictive Value of Cetuximab-Induced Skin Toxicity in Recurrent or Metastatic Squamous Cell Carcinoma of the Head and NECK . Front. Oncol. . 2018;**8**.https://www.frontiersin.org/articles/10.3389/fonc.2018.00616

282 Taylor CA, Watson RA, Tong O, *et al.* IL7 genetic variation and toxicity to immune checkpoint blockade in patients with melanoma. *Nat Med* 2022;**28**:2592–600. doi:10.1038/s41591-022-02095-5

283 Martins F, Sykiotis GP, Maillard M, *et al.* New therapeutic perspectives to manage refractory immune checkpoint-related toxicities. *Lancet Oncol* 2019;**20**:e54–64. doi:https://doi.org/10.1016/S1470-2045(18)30828-3

284 Pai C-CS, Simons DM, Lu X, *et al.* Tumor-conditional anti-CTLA4 uncouples antitumor efficacy from immunotherapy-related toxicity. *J Clin Invest* 2019;**129**:349–63. doi:10.1172/JCI123391

CHAPTER 8: APPENDICES

8.1 Appendix 1- Review of germline genetics and ICI toxicity

npj | genomic medicine

www.nature.com/npjgenmed

REVIEW ARTICLE OPEN

Check for updates

Germline genetic variation and predicting immune checkpoint inhibitor induced toxicity

Ik Shin Chin[1]✉, Aman Khan[1], Anna Olsson-Brown[2,3], Sophie Papa[4], Gary Middleton[5,6] and Claire Palles[1]

Immune checkpoint inhibitor (ICI) therapy has revolutionised the treatment of various cancer types. ICIs reinstate T-cell function to elicit an anti-cancer immune response. The resulting immune response can however have off-target effects which manifest as autoimmune type serious immune-related adverse events (irAE) in ~10–55% of patients treated. It is currently challenging to predict both who will experience irAEs and to what severity. Identification of patients at high risk of serious irAE would revolutionise patient care. While the pathogenesis driving irAE development is still unclear, host genetic factors are proposed to be key determinants of these events. This review presents current evidence supporting the role of the host genome in determining risk of irAE. We summarise the spectrum and timing of irAEs following treatment with ICIs and describe currently reported germline genetic variation associated with expression of immuno-modulatory factors within the cancer immunity cycle, development of autoimmune disease and irAE occurrence. We propose that germline genetic determinants of host immune function and autoimmune diseases could also explain risk of irAE development. We also endorse genome-wide association studies of patients being treated with ICIs to identify genetic variants that can be used in polygenic risk scores to predict risk of irAE.

npj Genomic Medicine (2022)7:73 ; https://doi.org/10.1038/s41525-022-00345-6

INTRODUCTION

Immune checkpoint inhibitors (ICI) are monoclonal antibodies that release the brakes off immune checkpoints such as cytotoxic T lymphocyte antigen (CTLA-4), programmed cell death 1 (PD-1), programmed cell death ligand 1 (PD-L1) and lymphocyte-activation gene 3 (Lag-3). They prevent the immune escape of cancer cells and ultimately cause cancer cell death[1]. They have revolutionised the management of several cancer types including melanoma, lung and urological cancers, which have historically had dismal prognoses in the metastatic setting. Commonly used ICIs include anti-PD-1 (nivolumab, pembrolizumab), anti-PD-L1 (atezolizumab, durvalumab, avelumab) and anti-CTLA-4 (ipilimumab) agents. Despite their notable successes in improving patients' survival, ICI-induced toxicities, also known as immune-related adverse events (irAEs), can be life-changing and in some cases fatal. As the indication for using ICIs expands and moves earlier in the treatment pathway in neoadjuvant and adjuvant settings, identifying patients at risk of developing irAEs is important to minimise the risk of serious toxicities while maximising the treatment benefit gained by patients.

There are currently no established predictive tools or biomarkers that can help detect patients at risk of irAEs. Clinicians mainly rely on patients' past clinical history including presence of autoimmune disease (AD) to provide a risk estimation. However, as patients with ADs have been excluded from most clinical trials of ICIs, it is hard to determine the effectiveness of this strategy. Given the estimated overall prevalence of autoimmune diseases of 7.6–9.4%, a high proportion of patients would miss out on potentially lifesaving cancer treatment were this to be used as a screening tool[2].

While the mechanism of irAEs is still unclear, host genetic factors are hypothesized to be key determinants. Associations between germline genetic polymorphisms and toxicity have already been established for several chemotherapeutic agents[3]. For example, variants in the dihydropyrimidine dehydrogenase *(DPYD)* gene have been shown to be associated with severe toxicities and are now screened for prior to administration of fluorouracil-based chemotherapy[3].

This review aims to summarise the clinical presentation of ICI-induced irAEs and the current evidence regarding germline factors and predisposition towards irAEs. As shown in Fig. 1, germline genetics may explain a significant proportion of the variation in patients' risk of irAEs. We hypothesise that risk is conferred via associated or intermediate phenotypes such as autoimmune diseases and levels of proteins involved in the cancer immunity cycle. Novel genetic determinants will only be uncovered by large-scale studies of common and rare variants in patients treated with ICIs.

THE SPECTRUM AND CHALLENGES OF IMMUNE-RELATED ADVERSE EVENTS

Given their similar presentation to autoimmune type events, irAEs are thought to be caused by an over-active immune response and disruption to immune homeostasis[1]. IrAEs can affect multiple organ systems including skin (manifesting as rash, pruritus, vitiligo), endocrine (thyroid, pituitary, adrenal disorders), gastro-intestinal (colitis), lung (pneumonitis), musculoskeletal (arthralgia, myalgia) and liver (hepatitis). Toxicities are graded based on the Common Terminology Criteria (CTCAE) for adverse events and are usually managed using oral, intravenous high-dose steroids, or immuno-modulatory agents. Early recognition and treatment of irAE is essential for symptom resolution. Some irAEs present similarly to chemotherapy or targeted therapy induced adverse events but their underlying cause and treatment approach may differ. For example, diarrhoea induced by ICIs is likely due to

[1]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK. [2]The Clatterbridge Cancer Centre, Wirral, UK. [3]The University of Liverpool, Liverpool, UK. [4]School of Cancer and Pharmaceutical Studies, King's College London, London, UK. [5]Institute of Immunology and Immunotherapy, University of Birmingham, Edgbaston, Birmingham, UK. [6]Queen Elizabeth Hospital, Birmingham, UK. ✉email: ikshinchin@nhs.net

Published in partnership with CEGMR, King Abdulaziz University

8.2 Appendix 2- CTCAE grading

CTCAE Grade	Description
1	Asymptomatic/mild symptoms. Clinical or diagnostic observations only. Intervention not needed
2	Moderate/minimal symptoms. Local or non-invasive intervention required. Limiting age-appropriate instrumental ADL
3	Severe/medically significant but not immediately life-threatening. Hospitalization or prolonged hospitalization needed. Disabling and limiting self-care ADL
4	Life-threatening and urgent intervention indicated.
5	Death

Appendix 2 Common Terminology Criteria for Adverse Events v5.0. Abbreviations: ADL- activities of daily living

8.3 Appendix 3- Immune-related genes

IL7	CD84	LINC00368	FSTL5	RNASET2	PLGRKT	LINC00824	PSORS1C1	APP	FKBP5
PNPT1	TNSF13B	RANBP3L	CLPB	ARID5B	CD274	SS18	RAB5C	CYYR1-AS1	TYK2
IFNG	APRIL	SLC1A3	LINC01537	AP000820.2	LINC00271	IRF1-AS1	ABO	RGMA	TG
LOC105377125	AKT	ABCC8	EMP2	C1QTNF6	ELMO1	RPL34P6	HLA-W	LINC02207	MAGI3
UNG	BLNK	LINC02240	MGAT5	AC002460.2	SH2D2A	GPR25	MICD	HLA	BANF2
IFNW1	STAT3	LINC02039	MIR3679	IL2RA	GSTM5P1	PDE10A	HLA-J	IL6	PDE8B
FUT2	NFAT	TRDN	C8orf37-AS1	RPL32P23	PPARG	SUOX	ZNRD1ASP	TNFA	TLR3
MIR146A	IRF4	SLC25A51P1	SMIM30	CFB	TNFRSF14	SBK1	PRR3	TLR2	RERE
CTLA-4	PRDM1	LOC102723883	PPP1R3A	CLEC2D	EDARADD	PANX1	ABCF1	MYD88	LINC01475
HLA	TNFRSF4	LINC01249	TSN	FLT3	ERBB3	IL2RA	MLN	PI3K	CCN4
RAC1	TNFSF4	LINC01248	LINC01826	SPATA13	CAMK4	ADCY7	LINC01016	NUAK1	TRAT1
GABRP	CD27	OPTC	LOC102546299	NCR3	TIMMDC1	AATF	ITPR3	MRPL48	PIK3R5
DSC2	CD70	ATP2B4	POM121L12	UQCRHP1	OL7R	DNASE1L3	RGS9	COL6A4P2	QRFPR
BAZ2B	GPR158	LRRC4C	LINC01446	MICA	RAD51B	IFITM2	IL1	WRN	TUBB4BP5
ZPLD1	ZNF701	MAT2B	RPS24	CLECL1	IRF3	TNFSF15	IL4	NRG1	RAB2A
MIR548AB	LINC01592	LINC02143	LINC00856	CD69	MACIR	ZMIZ1	IL13	LINC01541	JAK1
SEMA5A	SHISA9	NFATC2	LINC02497	TAP2	RN7SL714P	RNU6-320P	IFNA	LINC01899	BHLHE40-AS1
LOC105374704	ERCC4	ATP9A	LINC02501	TPO	CD226	LINC00892	CD40L	ZBTB14	GIGYF1

CDH6	PDE7B	KSR2	NEGR1	BTNL2	C1orf141	TMEM114	CDN	EPB41L3	FYN
OSBPL6	CRYBB2P1	RC3H1	LINC01360	TSBP1-AS1	IL23R	TRIM31	ATP	LINC00431	PLEKHA1
AGPS	GRK3	NBPF3	PLCXD3	MIR3681HG	ATG16L1	TRIM31-AS1	HMGB1	MYOC	MUC22
LOC102724355	EDIL3	FAM107B	OXCT1(	LINC02498	DAG1	SMARCA2	TLR	SGSM1	VANGL2
CFAP65	NBPF22P	CASC2	PPHLN1	RASGRP1	IGF2-AS	CD86	CD28	SHANK3	CASC19
LOC100129175	MAST3	FAM204A	XPO5	SLC1A2	INS-IGF2	IL21-AS1	CD80	CPEB2-DT	RPL13P2
RGMA	LINC00621	LINC00929	FSHR	MIR3188	SMAD3	CARD9	L12964	LINC02112	CD40
LINC01924	SGCG	GABRB3	NRXN1	RPL39P38	NKD1	NDUFA3P6	TNFSF9	FAM173B	LINC01550
ANKRD42	EAPP	PAQR3	NAMPT	CASC19	ADGRL2	RPL23AP12	IL2	IFT52	TSBP1-AS1
PACRG	ST6GAL2	LINC02064	CCDC71L	PCAT1	TENM3	HLA-S - AL645933.3	IL12	MYLK	HLA-DRA
PREX2	MIR548AU	LOC105374704	FBX04	ACAP1	LINC00993	HLA-DPA2	PDL1	RBFOX3	FBXL19-AS1
PTPM11	CSMD3	MIR4465	HCG27	TBX21	ANKRD30A	RNU6-1133P	PD1	LINC02233	DCHS2
SHP2	PDCD6IP	NMBR	HLA-C	HORMAD2	NKX2-3	C6orf15	CX3CL1	HLA-DQA1	VSIR
PDCD1	LOC101928135	KANK1	LINC01149	C12orf42	SLC25A28	CTHRC1P1	CXCL9	AL137856.1	LAG3
IFNL4	RGS6	DMRT1	HCP5	IFIH1	PVT1	KIF4CP	CXCL10	ATXN2	MICB
TAGAP	IFI44	TBX3	HLA-A	LINC02357	CX3CR1	CALM2P1	CCL5	SH2B3	VTCN1
CFAP36	ADGRL4	MED13L	HCG9	LINC02341	ANKRD55	AL671883.3	ICAM1	PTCSC2	HAVCR2
ANO5	TPK1	UCHL3	VEGF	LINC00243	INPP5B	HCG22	IDO	VAV3	
IL1RL1	CNTNAP2	LINC01146	IL10	MIR31HG	MIR4425	SLAMF6	TGFB	BACH2	
LINC02484	TYW3	LOC100288637	GIMAP5	H3P30	LINC02793	LINC01405	GITR/TNFRSF18	STAT4	

ARAP2	LHX8	HERC2P10	PTPN2	TCF7	CD247	LINC02265	BTLA	TSHR	
IL22RA1	CASC18	RNMT	SPACA6	RTKN2	LRRK2	FCRL3	HVEM/TNFRSF14	LPP	

Appendix 3 List of genes included in VEP filter

Selected genes associated with genes previously reported to be associated with irAE, immuno-modulatory traits within the cancer immunity cycle, response to ICI treatment, autoimmune diseases and from GWAS of serious irAE performed for the current dataset are listed here.

8.4 Appendix 4- Top genetic markers incorporated into the PRS of autoimmune diseases to predict irAE development

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
Ankylosing Spondylitis	HLA_B_27	**HLA_B**	rs148783236	**USP8**	1:200877850_T_G	**INAVA**
	HLA_B_2702	**HLA_B**	7:17197571_C_T	**AC073332.1, AHR**	rs4552569	**ZP3P1 - EDIL3**
	HLA_B_2705	**HLA_B**	22:45727565_T_G	**FAM118A**	2:62560332_A_G	**RN7SL51P - AC093159.2**
	16:28617885_G_A	**SULT1A1**	rs17095830	**ANO6**	rs4676410	**GPR35**
	rs10748781	**AL391684.1 - LINC01475**	rs17765610	**BACH2**	5:96121715_C_T	**ERAP1**
	rs10889676	**IL23R**	10:90749963_A_G	**AL157394.3, ACTA2**	2:102647300_G_A	**IL1R2 - AC007271.1**
	12:111862575_C_T	**SH2B3**	rs1801274	**FCGR2A**	rs56167332	**AC008691.1**
	rs11098964	**ANTXR2**	GSA-rs1860545	**TNFRSF1A**	16:3293407_T_C	**MEFV**
	1:25297184_G_A	**AL445471.2**	17:62147192_G_C	**AC005803.1, ERN1**	5:158818745_A_C	**AC008691.1**
	rs11616188	**LTBR - AC005840.2**	1:200960307_G_A	**KIF21B**	1:25305114_T_C	**AL445471.2 - MIR4425**
	14:88488821_T_C	**LINC01147**	6:20643727_G_A	**CDKAL1**	3:137441372_T_C	**AC117440.1 - NPM1P17**
	GSA-rs116488202	**AL671883.3 - FGFR3P1**	5:96174929_C_T	**AC009126.1**	rs6871626	**AC008691.1**
	1:200878727_G_A	**INAVA**	2:9402988_A_G	**ASAP2**	rs7115956	**RDX**
	rs1250550	**ZMIZ1**	rs27434	**ERAP1**	rs7282490	**GATD3A**
	10:81042475_G_A	**ZMIZ1**	5:96126308_A_G	**ERAP1**	11:128194450_G_T	**LINC02098 - ETS1**
	2:182048452_T_C	**LINC01934**	5:96252589_T_C	**AC009126.1, ERAP2**	rs8070463	**KPNB1 - TBKBP1**
	17:45690351_A_T	**NPEPPS, AC025682.2**	rs30187	**ERAP1**	2:102771282_T_A	**IL1R1**
	17:45575206_G_A	**MRPL45P2 - NPEPPS**	rs35164067	**CDC37**	GSA-rs9797244	**NOS2**
	1:154426264_C_T	**IL6R**	1:154426264_C_T	**IL6R**		

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
Coeliac disease	3:159637678_C_A	**IL12A-AS1**	rs1464510	**LPP**	3:46250374_T_C	**CCR3**
	2:103093243_T_C	**SLC9A4**	rs17035378	**PLEK**	1:2539400_T_C	**MMEL1**
	1:200881392_C_T	**INAVA**	rs1738074	**AL035530.2, AL035530.1, TAGAP**	2:61204917_T_G	**PUS10**

	rs10806425	**BACH2**	3:159647674_A_G	**IL12A-AS1**	rs4625	**DAG1**
	11:118726843_C_T	**NR**	rs17810546	**IL12A-AS1**	rs4675374	**ICOS**
	rs10903122	**AL445471.2 - MIR4425**	rs17849502	**NCF2, SMG7**	rs4819388	**ICOSLG**
	1:172674776_G_A	**NR**	6:159469574_A_G	**TAGAP-AS1**	6:128293562_C_T	**PTPRK**
	rs11221332	**ETS1**	3:188119901_G_A	**LPP**	7:37382465_T_G	**ELMO1**
	rs11712165	**ARHGAP31**	6:159470417_G_A	**TAGAP-AS1**	3:119123278_G_A	**ARHGAP31**
	1:200875242_T_A	**NR**	rs2187668	**HLA-DQA1**	1:2534978_G_A	**MMEL1**
	rs1250552	**ZMIZ1**	rs2327832	**BTF3L4P3 - AL356234.2**	1:61791863_G_A	**NFIA**
	2:182057640_T_G	**LINC01934**	12:6511996_C_A	**NR**	rs6806528	**FRMD4B**
	rs12928822	**RMI2**	18:12777573_C_G	**NR**	rs7426056	**CD28 - KRT18P39**
	rs13003464	**PUS10**	rs2816316	**AL390957.1**	4:123266521_A_T	**KIAA1109**
	rs13098911	**CCR3**	rs296547	**MROH3P**	rs7616215	**CCR3**
	rs13132308	**IL21-AS1**	rs3087243	**CTLA4 - ICOS**	rs802734	**THEMIS - PTPRK**
	3:33015469_G_T	**NR**	rs3184504	**SH2B3, ATXN2**	rs917997	**IL18RAP - SLC9A4**
	3:159623559_C_T	**NR**	6:128291199_C_T	**PTPRK**	22:37633851_C_T	**RAC2**
	1:192541472_C_G	**NR**	1:2526746_A_G	**MMEL1**	14:69271969_C_A	**NR**
	2:103086770_A_C	**NR**	rs9792269	**RN7SKP226 - AC100822.1**		

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
Crohn's disease	rs10045431	**AC008691.1**	rs17221417	**NOD2**	rs4780355	**RMI2**
	rs10051722	**AC006525.1**	rs17293632	**SMAD3**	22:37258503_T_C	**NCF4, NCF4-AS1**
	rs10181042	**PUS10**	rs17309827	**SLC22A23**	1:112262826_A_T	**INKA2**
	rs10185424	**AC007271.1**	rs1736020	**AJ009632.2**	rs4902642	**AL132986.1 - ZFP36L1**
	rs102275	**TMEM258**	rs1736135	**AJ009632.2**	rs4976646	**RGS14**
	rs10486483	**SKAP2**	rs17391694	**GIPC2 - RNFT1P2**	rs516246	**FUT2**
	16:82867456_C_A	**CDH13**	rs174537	**TMEM258, MYRF**	3:188401160_C_T	**LPP**
	rs10495903	**THADA**	rs17582416	**LINC02635 - CUL2**	rs568617	**FIBP**
	rs10734105	**TCERG1L - LINC01164**	rs1799964	**LTA - TNF**	rs5763767	**HORMAD2**

rs10761659	**AC067751.1**	rs1801274	**FCGR2A**	rs6074022	**RPL13P2 - CD40**
rs10781499	**CARD9**	rs181359	**UBE2L3**	rs6478106	**AL390240.1 - TNFSF15**
1:186875459_G_T	**PLA2G4A**	rs1819658	**AC026884.1 - IPMK**	rs6545946	**AC093159.1 - RN7SL18P**
rs10865331	**RN7SL51P - AC093159.2**	rs1842076	**KRT18P56 - LINC00604**	rs6556412	**AC008691.1**
rs10947261	**BTNL2, TSBP1-AS1**	rs1847472	**BACH2**	rs6568421	**RN7SKP211 - PRDM1**
rs11010067	**LINC02635 - CUL2**	rs1998598	**AL365258.2, DENND1B**	rs6651252	**LINC00824**
rs11167764	**AC005740.2 - NDFIP1**	rs2024092	**SBNO2**	rs6679677	**PHTF1 - RSBN1**
rs11175593	**LRRK2-DT, LRRK2, LINC02471**	rs2058660	**IL18RAP**	rs670523	**RIT1**
rs11209026	**IL23R**	13:43052880_G_A	**LINC02341**	rs6708413	**IL18RAP**
rs11229030	**SLC43A3 - RNA5SP341**	16:50756540_G_C	**NOD2**	2:198896895_A_G	**PLCL1**
11:72863697_A_G	**FCHSD2 - AP002761.2**	rs2066847	**NR**	rs6740462	**AC007389.1, AC012370.2**
rs11465804	**IL23R**	rs2076756	**NOD2**	rs6856616	**AC098680.1 - LINC02513**
6:31269693_T_C	**HLA-B**	rs212388	**AL035530.1, AL035530.2, AL356417.3**	rs6863411	**NDFIP1**
rs11564258	**MUC19, AC107023.1**	rs2149085	**RNASET2 - Z94721.1**	rs6908425	**CDKAL1**
rs11584383	**MROH3P - KIF21B**	rs2155219	**EMSY - AP001189.2**	rs694739	**AP003774.2 - CCDC88B**
rs11641184	**LITAF**	rs2227551	**PLAU**	8:90875918_T_C	**AF117829.2 - RNU6-925P**
rs11681525	**TEX41**	rs224090	**AC067751.1**	rs7076156	**AC067752.1, AC024598.1**
rs11741861	**IRGM, ZNF300**	10:64477836_G_A	**AC067751.1**	rs7097656	**TSPAN14**
rs11742570	**AC108105.1 - AC093277.1**	rs2274910	**ITLN1**	rs713875	**AC002378.1**
rs11743851	**CDC42SE2**	rs2284553	**IFNGR2**	rs7236492	**AC018445.5, NFATC1**
rs11747270	**IRGM**	rs2301436	**FGFR1OP, Z94721.2**	rs727563	**ACO2**
17:40570772_A_C	**CAVIN1**	rs2361755	**AC092723.4**	rs7282490	**GATD3A**
rs11879191	**CDC37**	rs2382817	**TMBIM1, PNKD**	13:41558110_A_G	**ELF1**
2:5664008_G_T	**AC073143.1 - AC107057.1**	rs2413583	**AL031590.1**	rs736289	**SLC7A10 - AC008738.6**
rs12242110	**RNU6-794P - CCNY**	1:197631141_G_A	**DENND1B**	rs7423615	**SP140**
10:81032532_A_G	**ZMIZ1**	rs2549794	**ERAP2, AC009126.1**	4:48363245_A_G	**SLAIN2**
rs1250550	**ZMIZ1**	3:53133149_C_G	**AC096887.1, RFT1**	rs744166	**STAT3**
rs12521868	**IRF1-AS1, AC116366.3**	rs2641348	**ADAM30**	rs751728	**MLN**
16:28340945_A_G	**SBK1 - AC138904.2**	rs2797685	**PER3, Z98884.1**	rs7517847	**IL23R, C1orf141**

	rs12677663	**SBSPON**	rs281379	**FUT2 - MAMSTR**	rs7554511	**INAVA**
	rs12718244	**SPATA48**	rs2838519	**GATD3A**	rs7555082	**AL157402.1 - PTPRC**
	rs12720356	**TYK2**	rs2930047	**DAP**	rs762421	**GATD3A**
	rs12722489	**IL2RA**	6:83648056_C_T	**UBE3D, AL355613.1**	1:67648596_G_A	**C1orf141, IL23R**
	rs12722515	**IL2RA**	rs2945412	**KSR1**	rs7702331	**AC116345.1**
	rs1292053	**TUBD1**	rs3024505	**AL591846.1 - IL10**	5:101946798_A_G	**LINC00491, LINC00492**
	rs12942547	**STAT3**	rs3091315	**CCL2 - CCL7**	5:150270420_A_G	**IRGM**
	rs13003464	**PUS10**	5:131402738_C_T	**IL3 - CSF2**	rs7746082	**RN7SKP211 - PRDM1**
	rs13073817	**SATB1-AS1**	rs1142287	**SCAMP3**	rs7758080	**TAB2-AS1, TAB2**
	rs13204048	**SLC22A23**	1:159799910_G_A	**SLAMF8**	rs7765379	**MTCO3P1 - AL662789.1**
	rs13300218	**NOTCH1**	rs359457	**LINC01485 - CPEB4**	rs7773324	**DUSP22 - IRF4**
	rs13407913	**ADCY3**	rs3742130	**UBAC2, GPR18**	rs780093	**GCKR**
	rs13428812	**DNMT3A**	2:241570249_A_C	**GPR35**	rs7954567	**LTBR**
	rs1363907	**AC009126.1, ERAP2**	rs3764147	**LACC1**	14:88472595_C_T	**GPR65**
	6:32659443_A_C	**HLA-DQB1 - MTCO3P1**	rs3851228	**TRAF3IP2-AS1**	13:42917861_T_C	**LINC02341**
	rs1456893	**AC020743.2**	rs3853824	**C17orf67**	rs921720	**AC091114.1**
	rs1456896	**AC020743.2 - AC020743.3**	rs395157	**OSMR**	rs9258260	**AL645939.6**
	rs1487630	**AC098680.1 - LINC02513**	rs4077515	**CARD9**	rs9264942	**HLA-B**
	rs151181	**CLN3, AC138894.1**	rs415890	**Z94721.1**	rs9271366	**HLA-DRB1 - HLA-DQA1**
	rs1517352	**STAT4**	rs4256159	**SATB1-AS1**	rs9286879	**AL031599.1**
	rs1569328	**AF111167.1 - FOS**	rs4703855	**AC093278.1 - YBX1P5**	rs9525625	**LINC02341**
	rs17119	**AL138720.1**	rs4768236	**LRRK2**	rs9891119	**STAT3**

AD trait	**SNP**	**Mapped gene**
Addison's disease	rs3998178	**HLA-DQA1 - HLA-DQB1**
	rs10806425	**BACH2**
	GSA-rs74203920	**AIRE**

	rs2476601	**AL137856.1, PTPN22**
	rs2075876	**AIRE**
	rs1464510	**LPP**
	2:204707138_G_A	**RNU6-474P - CTLA4**
	rs11203203	**UBASH3A**
	19:52204248_G_A	**SPACA6**
	12:111932800_C_T	**ATXN2**

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
Type 1 diabetes	rs10509540	**AL353149.1 - RNLS**	rs2281808	**SIRPG**	rs516246	**FUT2**
	rs10517086	**LINC02357**	rs2290400	**GSDMB**	14:101306447_T_C	**AL117190.2, MEG3, AL117190.1**
	GSA-rs11170445	**HIGD1AP1 - EIF4A1P4**	rs2292239	**ERBB3**	rs5753037	**AC002378.1**
	rs11171739	**ERBB3**	22:37587111_A_C	**C1QTNF6**	rs61839660	**IL2RA**
	21:43825357_C_G	**UBASH3A**	rs2476601	**AL137856.1, PTPN22**	7:50465830_A_G	**AC124014.1, IKZF1**
	rs113010081	**CCRL2 - LINC02009**	4:166574267_A_G	**AC080079.2**	rs653178	**ATXN2**
	5:35883251_C_G	**IL7R - AC112204.3**	rs2664170	**NR**	1:200814959_T_C	**CAMSAP2**
	16:11194771_C_T	**CLEC16A**	rs3024505	**AL591846.1 - IL10**	rs7020673	**GLIS3**
	rs1456988	**LINC01550 - AL163932.1**	rs3087243	**CTLA4 - ICOS**	12:56384804_G_A	**RAB5B**
	rs1465788	**ZFP36L1 - MAGOH3P**	19:10463118_G_C	**TYK2**	12:56435504_G_C	**RPS26, AC034102.3**
	16:28505660_G_C	**CLN3**	rs3764021	**CLEC2D**	rs7202877	**CTRB2 - CTRB1**
	6:126752884_C_G	**AL356534.1**	10:8102272_G_A	**GATA3**	rs7221109	**AC004585.2 - SMARCE1**
	18:67526644_T_C	**CD226**	11:2184848_G_T	**INS - TH**	rs72727394	**RASGRP1**
	rs1701704	**IKZF4**	19:47219122_T_C	**PRKD2**	rs72928038	**BACH2**
	6:32627833_A_G	**HLA-DQB1**	rs425105	**PRKD2**	rs75793288	**KIAA1109**
	rs1893217	**PTPN2**	6:26423560_C_T	**BTN2A3P**	rs763361	**CD226**
	2:142300935_T_G	**LRP1B**	rs4763879	**CD69**	rs7804356	**SKAP2**
	rs193778	**RMI2**	rs478222	**EFR3B**	16:75252327_G_A	**CTRB2 - CTRB1**

	rs1990760	**IFIH1**	rs4788084	**IL27 - NUPR1**	rs924043	**AL049612.1 - AL603783.1**
	rs2111485	**FAP - IFIH1**	2:111615079_T_G	**ACOXL**	rs9388489	**AL356534.1**
	21:22096544_T_A	**LINC00320**	rs4900384	**LINC01550 - AL163932.1**	14:101306045_A_G	**AL117190.2, AL117190.1, MEG3**
	1:64107284_C_T	**PGM1**	rs4948088	**AC004830.2**	13:100081766_C_T	**AL136961.1**

AD trait	SNP	Mapped gene
Myasthenia gravis	6:32517045_G_C	**HLA-DRB5 - RNU1-61P**
	6:31515799_G_A	**ATP6V1G2, NFKBIL1**
	HLA_DQA1_0501	**HLA-DQA1**
	6:29947740_T_C	**HCG9 - DDX39BP2**
	rs231770	**RNU6-474P - CTLA4**
	HLA_DQA1_05	**HLA-DQA1**
	8:81364205_G_A	**AC034114.2**
	rs4263037	**TNFRSF11A**

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
Psoriasis	rs4406273	**HLA-B**	12:112241766_G_A	**ALDH2**	rs10782001	**FBXL19**
	rs13200483	**NR**	16:31021078_C_T	**STX1B**	15:31637666_T_C	**KLF13, KLF13**
	rs10484554	**HLA-B**	16:11354091_A_G	**RMI2**	9:110792282_T_C	**NR**
	rs12188300	**NR**	10:75601596_C_T	**CAMK2G**	11:128406438_A_G	**ETS1**
	GSA-rs35960711	**SCRN1**	rs20541	**IL13, TH2LCRR**	rs694739	**AP003774.2 - CCDC88B**
	rs33980500	**TRAF3IP2-AS1, TRAF3IP2**	11:128391937_C_T	**ETS1**	10:102038641_G_A	**BLOC1S2**
	19:10463118_G_C	**NR**	rs2675662	**CAMK2G**	rs6590334	**ETS1**
	rs7709212	**AC008691.1**	20:48556229_C_T	**RNF114**	1:67670755_C_T	**C1orf141, IL23R**
	rs2233278	**TNIP1**	3:16996623_G_A	**PLCL2**	20:48522330_G_A	**SPATA2**
	6:111929862_G_T	**NR**	3:16996035_G_A	**PLCL2, PLCL2**	rs10789285	**LINC02791 - LINC01758**

rs13437088	**HLA-S - AL645933.3**	10:81032885_G_A	**ZMIZ1**	7:37386237_C_A	**ELMO1**
5:158804928_G_T	**AC008691.1**	2:62560332_A_G	**RN7SL51P - AC093159.2**	12:10597207_C_T	**KLRC1**
rs17728338	**TNIP1 - ANXA6**	exm-rs492602	**FUT2**	rs12586317	**PRORP, AL121594.1**
rs2778031	**AL451142.2 - RPSAP49**	22:21974703_G_C	**UBE2L3**	rs34517439	**DNAJB4, GIPC2**
rs2546890	**AC008691.1**	18:51816394_C_T	**POLI**	17:40561579_G_C	**CAVIN1**
GSA-rs6677595	**LCE3B - LCE3A**	rs1581803	**LCE3B - LCE3A**	rs5855282	**NR**
rs4085613	**LCE3E - LCE3D**	1:25293084_T_C	**RUNX3 - AL445471.2**	rs112768831	**NR**
rs2082412	**UBLCP1 - AC008691.1**	1:152592184_C_T	**LCE3B - LCE3A**	11:65593444_A_T	**CFL1**
1:152591953_G_C	**NR**	rs2066807	**STAT2**	14:98668778_AAG_A	**NR**
2:61072183_T_A	**REL-DT**	rs149442660	**GSTCD, INTS12**	17:73890363_G_A	**TRIM65**
rs582757	**TNFAIP3**	rs465969	**MFSD4B, REV3L**	6:108370585_T_C	**OSTM1**
12:56741228_G_A	**STAT2**	1:24518643_A_G	**IFNLR1 - LINC02800**	rs1990760	**IFIH1**
2:163167746_C_G	**IFIH1**	6:577820_A_G	**EXOC2**	13:45321731_C_A	**NR**
5:131996445_A_G	**IL13**	10:81043743_A_G	**ZMIZ1**	11:64053157_T_G	**GPR137**
17:26124908_G_A	**NOS2**	rs702873	**LINC01185**	17:78175483_T_C	**CARD14**
20:48574454_G_A	**KRT18P4**	5:96101944_A_G	**CAST, ERAP1**	10:64369999_C_G	**AC024598.1, AC067752.1**
5:96118852_G_C	**ERAP1**	10:75653872_C_T	**CAMK2G - PLAU**	12:112059557_C_T	**ATXN2-AS - BRAP**
5:96119273_A_G	**ERAP1**	rs12720356	**TYK2**	12:122668326_G_A	**LRRC43**
rs240993	**MFSD4B, REV3L**	19:49129346_C_T	**SPHK2**	1:206655331_C_T	**IKBKE**
rs2145623	**NR**	6:20678430_G_C	**CDKAL1**	1:8268095_G_A	**LINC01714**
4:122754459_G_T	**BBS7**	rs9533962	**AL356515.1 - LINC00330**	9:110817020_A_G	**AC068050.1 - CHCHD4P2**
1:25291010_A_T	**RUNX3**	rs7637230	**LINC02085**	16:31000809_G_A	**STX1B**
rs7748720	**CDKAL1**	rs842625	**LINC01185**	17:65870073_C_T	**BPTF**
rs17716942	**KCNH7**	7:37385365_A_G	**ELMO1**	10:89824771_A_G	**MED6P1 - AL353149.1**
2:61083506_T_C	**LINC01185**	17:40536396_T_C	**STAT3**	rs11652075	**CARD14, AC087741.1**
GSA-rs892085	**QTRT1**	rs4804528	**DNM2**	rs2201841	**C1orf141, IL23R**
rs610604	**TNFAIP3**	1:172675097_T_C	**FASLG - SLC25A38P1**	rs34394770	**COG6**
12:56750204_C_T	**STAT2**	rs7769061	**TRAF3IP2**	rs12650590	**AC107208.1 - GUCY1A1**
1:152603842_C_A	**LCE3A - LCEP4**	1:8286009_G_A	**NR**	13:99950260_G_A	**GPR183, UBAC2**

	16:31004812_T_C	**STX1B**	18:51819750_C_T	**POLI**	19:49206631_A_T	**FUT2**
	rs10794648	**NR**	rs2853694	**AC008691.1, IL12B**	22:21979289_T_C	**UBE2L3 - YDJC**
	11:109962432_A_G	**NR**	rs2066808	**STAT2**	rs8128234	**RUNX1**
					3:189615475_C_A	**TP63 - P3H2**

AD trait	SNP	Mapped gene
Steven Johnson's Syndrome	rs4917014	**AC020743.2 - AC020743.3**
	rs9469003	**AL645933.5**
	rs3130501	**POU5F1**
	15:73729483_G_C	**HCN4 - REC114**
	rs3094188	**POU5F1, PSORS1C3**

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
Ulcerative colitis	6:32612397_A_G	**HLA-DQA1**	rs17229285	**AC019330.1**	rs113010081	**CCRL2 - LINC02009**
	6:31213225_A_G	**NR**	rs4151657	**CFB, CFB, AL645922.1**	rs12103	**INTS11**
	rs6426833	**AL391883.1 - OTUD3**	rs3766606	**PARK7**	rs1517352	**STAT4**
	rs6017342	**LINC01620**	rs941823	**LINC00598**	9:5260079_C_T	**NR**
	rs3024505	**AL591846.1 - IL10**	HLA_DRB1_0103	**HLA-DRB1**	rs7210086	**AC080037.2 - SLC39A11**
	rs1801274	**FCGR2A**	rs17780256	**SLC39A11**	rs1126510	**PTGIR**
	rs3806308	**AL391883.1**	1:2501338_C_T	**NR**	rs4656958	**ITLN1 - AL354714.2**
	10:101284237_T_G	**NR**	rs3774959	**NFKB1**	rs4976646	**RGS14**
	rs7554511	**INAVA**	rs17694108	**SLC7A10 - AC008738.6**	rs4743820	**AL158071.3, LINC00484, AL158071.1**
	7:107480315_T_C	**NR**	rs12720356	**TYK2**	16:28340945_A_G	**NR**
	rs6920220	**AL356234.2**	rs2838519	**GATD3A**	rs1077773	**AC019117.4, AC019117.3**
	rs4654925	**OTUD3**	rs4676406	**GPR35 - AC011298.3**	rs10889677	**IL23R**
	rs10761659	**AC067751.1**	1:67648596_G_A	**C1orf141, IL23R**	rs9847710	**SFMBT1, AC096887.1**
	rs561722	**NXPE2P1 - AC020549.3**	7:98760504_G_A	**NR**	rs11229555	**GLYAT**
	rs2816958	**NR5A2**	rs2189234	**TET2**	20:62348907_A_G	**ZGPAT, RP4-583P15.15**
	rs4845604	**RORC**	rs13300218	**NOTCH1**	rs913678	**LINC01271 - RN7SL636P**

rs798502	**AMZ1, GNA12**	rs6667605	**AL139246.2 - AL139246.3**	rs2382817	**TMBIM1, PNKD**
rs17207986	**AL662884.2, AL662884.5, TNXB**	rs483905	**MAML2**	rs11739663	**MIR4456 - AC106772.1**
rs17085007	**FGFR1OP2P1 - AL160035.1**	rs11150589	**AC116348.1 - AC116348.2**	rs4722672	**AC004080.2**
rs4246905	**TNFSF15**	rs7240004	**CTIF - SMAD7**	rs11010067	**LINC02635 - CUL2**
rs3749946	**AL645933.4**	rs254560	**C5orf66**	rs28374715	**CHP1**
rs7282490	**GATD3A**	rs16940186	**AC092723.4**	rs6088765	**MMP24OS, PROCR**
rs1182188	**GNA12**	rs11641184	**LITAF**	rs1016883	**PLCL1**
rs10185424	**AC007271.1**	4:123161619_A_G	**KIAA1109**	rs1728785	**ZFP90**
rs3851228	**TRAF3IP2-AS1**	rs1405108	**AC019330.1 - AC020718.1**	rs11083840	**CALM3 - PTGIR**
rs4728142	**AC025594.1 - IRF5**	rs549182	**NOTCH4 - TSBP1-AS1**	rs6499188	**CDH3**
rs17736589	**CYTH1**				

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
Rheumatoid arthritis	rs660895	**HLA-DRB1 - HLA-DQA1**	rs4409785	**AP000820.2**	2:65580221_T_C	**SPRED2**
	rs2476601	**AL137856.1, PTPN22**	rs3093023	**AL121935.1, CCR6, Z94721.2**	rs2072438	**PHF19**
	rs13192471	**HLA-DQB1 - MTCO3P1**	rs2561477	**MACIR**	rs67250450	**JAZF1**
	6:138227364_T_G	**AL591468.1 - LINC02528**	6:138003822_G_C	**AL356234.2**	10:8104722_C_T	**GATA3**
	6:138230389_G_A	**AL591468.1 - LINC02528**	rs2451258	**AL035530.2, AL035530.1, AL356417.3**	rs227163	**UTS2 - TNFRSF9**
	rs7731626	**ANKRD55**	rs1516971	**LINC00824, LINC00824**	rs2671692	**WDFY4, AC060234.3**
	ilmnseq_rs2301888_ilmndup1	**PADI4, PADI4**	rs2812378	**AL162231.5**	4:10727357_T_C	**AC084048.1 - LINC02498**
	rs2233424	**NFKBIE - TMEM151B**	rs3806624	**EOMES - LINC01980**	rs12795702	**LINC02098 - ETS1**
	rs71508903	**ARID5B**	rs4452313	**PLCL2, PLCL2**	12:111933545_C_A	**ATXN2**
	rs3087243	**CTLA4 - ICOS**	rs10985070	**PHF19**	16:11833886_G_T	**TXNDC11**
	rs8032939	**RASGRP1**	2:100764004_T_G	**AFF3**	10:31411355_C_T	**AL359546.1**
	rs12617656	**DPP4**	21:36738242_G_A	**RUNX1**	6:106676470_C_T	**ATG5**
	7:74026152_AC_A	**NR**	6:382537_G_A	**DUSP22 - IRF4**	10:6122009_T_A	**RPL32P23 - RBM17**
	22:39747780_G_A	**SYNGR1**	1:114388804_T_C	**PTPN22, AL137856.1**	15:70048116_A_T	**DRAIC**

13:40355913_T_C	**COG6**	1:154425456_G_A	**IL6R**	rs2075876	**AIRE**
GSA-rs2841275	**AL583810.3 - PLD4**	rs3761959	**FCRL3**	rs7579944	**H3P5 - LBH**
rs8026898	**DRAIC**	11:72404893_C_T	**ARAP1, AP003065.2**	7:28166442_G_A	**JAZF1**
rs2233434	**NFKBIE**	9:123668199_G_T	**TRAF1**	14:35861800_C_T	**AL133163.1 - NFKBIA**
rs3093024	**Z94721.2, CCR6, AL121935.1**	22:21916361_G_A	**UBE2L3**	rs3125734	**RTKN2**
2:100760172_T_C	**AFF3**	10:6108340_A_G	**IL2RA - RPL32P23**	rs4272	**CDK6**
rs6920220	**AL356234.2**	15:91038883_C_CT	**IQGAP1**	1:17674402_C_T	**PADI4, PADI4**
rs9653442	**LINC01104**	GSA-rs4602367	**PLCL2, PLCL2**	21:26769507_T_C	**LINC00158**
rs4239702	**CD40**	1:17413899_T_A	**PADI2**	17:4503195_G_A	**SMTNL2**
15:70010647_C_T	**DRAIC**	20:62484008_C_T	**C20orf181 - AL118506.1**	rs6705628	**DGUOK-AS1**
18:12881361_A_G	**PTPN2**	11:9752741_C_T	**SWAP70**	12:111833788_G_A	**LINC02356**
11:118729391_A_G	**AP004609.1 - CXCR5**	rs805297	**APOM**	rs678347	**AP001208.1**
4:26111593_G_A	**LINC02357 - AC097714.1**	rs951005	**AL162231.5**	rs998731	**TPD52**
rs12379034	**CDK5RAP2**	21:45653764_G_C	**ICOSLG**	10:63959929_C_T	**RTKN2**
6:32448599_C_T	**HLA-DRB9 - HLA-DRB5**	rs9979383	**RUNX1**	rs6732565	**ACOXL**
rs34695944	**REL**	rs934734	**SPRED2**	rs4780401	**AC007613.1 - ZC3H7A**
6:32451788_G_C	**HLA-DRB9 - HLA-DRB5**	11:9773517_C_A	**SWAP70**	rs726288	**SFTPD, AL512662.2**
10:6100725_G_A	**IL2RA**	10:6517167_T_C	**PRKCQ**	rs10931468	**NAB1**
rs11574914	**AL162231.5**	5:133423890_T_G	**AC008608.1 - TCF7**	1:2534978_G_A	**MMEL1, MMEL1**
rs1858037	**SPRED2**	rs3781913	**PDE2A**	2:100672692_C_T	**AFF3**
10:6098949_C_T	**IL2RA**	rs4936059	**AP001122.1**	rs42034	**CDK6**
rs1854853	**AL121935.1, CCR6, Z94721.2**	rs9372120	**ATG5**	1:157674997_G_T	**FCRL3 - AL356276.2**
rs3761847	**TRAF1**	rs624988	**GAPDHP64 - NEFHP1**	rs1877030	**AC009283.1 - NEUROD2**
4:10726520_C_T	**AC084048.1 - LINC02498**	19:10771941_C_T	**ILF3**	rs73081554	**RPP14, HTD2**
1:2523811_G_A	**MMEL1, MMEL1**	rs2469434	**CD226**	10:64063077_T_C	**LINC02621**
11:118693161_T_A	**AP004609.4 - SETP16**	3:136402060_G_A	**STAG1**	10:8102272_G_A	**GATA3**
1:2483961_G_A	**TNFRSF14-AS1, TNFRSF14-AS1**	11:64042997_G_A	**BAD, GPR137**	16:23871206_A_AAGGGAG	**PRKCB**
rs1950897	**RAD51B**	rs11089637	**UBE2L3 - YDJC**	1:186632317_C_T	**AL049198.1 - PTGS2**
rs1893592	**UBASH3A**	rs3781094	**GATA3**	22:45746152_A_AT	**SMC1B**

17:38031857_T_G	**ZPBP2**	rs10175798	**H3P5 - LBH**	rs883220	**AL390839.1**
6:159489791_G_A	**AL356417.3, AL035530.2, AL035530.1**	10:31415106_C_T	**AL359546.1**	rs12529514	**AL022396.1 - CD83**
1:173349725_A_C	**AL645568.2 - AL645568.1**	rs1748041	**PADI4, PADI4**	rs968567	**FADS2, FADS1**
1:160417157_T_C	**VANGL2 - SLAMF6**	21:34775444_T_C	**IFNGR2, IFNGR2**	11:60909581_C_A	**VPS37C**
TOP2-rs7097397	**WDFY4**	1:173166247_A_G	**TNFSF4**	16:30626616_C_T	**AC093249.3, ZNF689**
rs13330176	**AC092723.4 - AC092723.3**	4:78508197_T_G	**CXCL13**	14:69213851_G_A	**AL132986.1 - ZFP36L1**
rs13017599	**NONOP2**	1:17429774_G_T	**PADI2**	rs911263	**RAD51B**
12:56469185_C_T	**RPS26 - ERBB3**	6:138289848_T_C	**LINC02865**	3:27777779_G_A	**EOMES - LINC01980**
rs28411352	**MTF1**	6:36355654_A_C	**ETV7-AS1**	21:34764288_C_A	**IFNAR1 - IFNGR2**
9:123662618_T_C	**PHF19 - TRAF1**	rs72634030	**RABEP1, AC015727.1, NUP88**	rs2664035	**TEC**
rs4728142	**AC025594.1 - IRF5**	14:61940675_G_A	**PRKCH**	rs26232	**MACIR**
14:69260588_C_T	**ZFP36L1**	rs6715284	**FLACC1**	rs4963581	**AC087312.1 - AC023796.1**
2:30445322_A_G	**H3P5 - LBH**	1:38633879_A_T	**AL390839.1 - AL390839.2**	11:68859848_C_CGGCTAAGGGGAGAGAGCCCGGAAGCCACGCACACCGGCGCAAGGGGCGGATA	**NR**
1:2528133_T_C	**MMEL1, MMEL1**	10:8095340_G_A	**GATA3, GATA3-AS1**	10:64036881_G_C	**RTKN2 - LINC02621**
2:204636190_A_G	**KRT18P39 - NPM1P33**	rs2228145.1	**IL6R**	rs13315591	**FAM107A**
6:14107197_C_T	**AL022396.1 - CD83**	1:173299743_C_T	**TNFSF4 - AL645568.3**	5:10704797_T_C	**DAP**
rs12413578	**AC044784.1 - LINC00709**				

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
SLE	HLA_B_0801	**HLA-B**	7:75209951_G_C	**HIP1**	1:173347837_G_A	**AL645568.2 - AL645568.1**
	HLA_DQA1_0102	**HLA-DQA1**	9:102585545_T_C	**NR4A3, STX17-AS1**	rs4917014	**AC020743.2 - AC020743.3**
	HLA_DQA1_0501	**HLA-DQA1**	7:73566677_G_A	**LIMK1 - EIF4H**	10:105003721_T_G	**ST13P13 - RPEL1**
	HLA_DQB1_0201	**HLA-DQB1**	1:192543837_A_T	**AL390957.1**	rs494003	**AP5B1**
	HLA_DQB1_0301	**HLA-DQB1**	7:28189411_T_C	**JAZF1**	rs4948496	**ARID5B**
	HLA_DQB1_0602	**HLA-DQB1**	2:30479857_C_A	**LBH**	rs4963128	**PHRF1, PHRF1**
	HLA_DRB1_0301	**HLA-DRB1**	rs1734787	**MECP2**	6:138195151_G_T	**TNFAIP3**

	HLA_DRB1_1501	**HLA-DRB1**	rs17603856	**ATXN1**	rs548234	**ATG5**
	4:184609373_C_T	**TRAPPC11**	1:246444082_C_T	**SMYD3**	5:100184647_C_A	**ST8SIA4**
	rs10036748	**TNIP1**	1:183542323_C_T	**NCF2, SMG7**	rs56154925	**AC010327.5 - TMEM86B**
	2:213890232_G_T	**IKZF2**	8:42033519_G_A	**PLAT**	3:159728987_C_T	**IL12A-AS1**
	1:117076547_G_T	**CD58, NAP1L4P1**	3:58318464_A_C	**AC098479.1 - PXK**	3:159729059_G_A	**IL12A-AS1**
	7:28142186_G_T	**JAZF1**	13:100091300_A_G	**AL136961.1 - TM9SF2**	11:625085_A_G	**CDHR5, CDHR5**
	19:55739813_G_A	**AC010327.6, TMEM86B**	9:4984530_G_C	**JAK2**	GSA-rs58721818	**AL591468.1 - LINC02528**
	rs10488631	**AC025594.2 - TNPO3**	rs1913517	**WDFY4, LRRC18, AC060234.3**	6:91002494_A_G	**BACH2**
	rs10499197	**LINC02539 - WAKMAR2**	7:73537902_T_C	**LIMK1 - EIF4H**	12:111973358_A_G	**ATXN2**
	12:129278864_A_G	**SLC15A4**	rs1990760	**IFIH1**	20:1541752_G_T	**AL049634.2**
	1:173346343_C_T	**AL645568.2 - AL645568.1**	rs2176082	**PXK**	rs6445975	**PXK**
	rs10774625	**ATXN2**	rs2205960	**TNFSF4 - AL645568.3**	rs653178	**ATXN2**
	rs10845606	**AC008115.2, GPR19**	rs2230926	**TNFAIP3**	6:106588806_A_C	**ATG5**
	rs10911628	**AL713852.1 - EDEM3**	rs2254546	**AF131216.5 - BLK**	rs6590330	**LINC02098 - ETS1**
	1:173251856_A_G	**TNFSF4 - AL645568.3**	rs2275247	**KIAA0319L**	1:67802371_A_C	**IL12RB2**
	rs10946940	**CD83P1 - RNU6-471P**	17:4712617_C_T	**PLD2**	1:161478810_G_A	**FCGR2A**
	7:128589633_C_T	**IRF5**	rs2289583	**SCAMP5**	rs6679677	**PHTF1 - RSBN1**
	10:73501066_C_A	**CDH23**	1:206643772_C_G	**SRGAP2 - IKBKE**	rs6705628	**DGUOK-AS1**
	12:129289190_G_A	**SLC15A4**	6:106667994_C_T	**ATG5**	3:188470238_C_T	**LPP**
	12:129294333_T_C	**SLC15A4**	rs2301271	**HLA-DQB2**	rs6804441	**CD80**
	rs11066301	**PTPN11**	rs2304256	**TYK2**	6:138242437_T_C	**AL591468.1 - LINC02528**
	rs11073328	**FAM98B**	rs2327832	**BTF3L4P3 - AL356234.2**	rs704840	**TNFSF4 - AL645568.3**
	19:10462513_C_T	**TYK2**	2:144020974_C_T	**ARHGAP15**	rs7197475	**AC093249.1 - PRR14**
	19:10466123_C_T	**TYK2**	19:18386634_A_C	**IQCN - JUND**	16:11177801_A_G	**CLEC16A**
	16:86015316_A_G	**AC092723.4**	1:173337747_G_A	**AL645568.2 - AL645568.1**	rs729302	**AC025594.1 - IRF5**
	7:50299077_A_G	**AC020743.2**	5:159887336_A_G	**MIR3142HG**	rs7300146	**GLT1D1**
	11:617228_G_A	**CDHR5, CDHR5**	rs2431697	**MIR3142HG**	rs73135369	**GTF2IRD1**
	11:72863697_A_G	**FCHSD2 - AP002761.2**	8:128197088_A_G	**CASC19, PCAT1**	13:41558110_A_G	**ELF1**
	1:157497160_A_G	**FCRL5**	GSA-rs268134	**SPRED2**	rs74290525	**SLC44A4**

	3:160312921_G_T	**RPL6P8 - ARL14**	11:35088399_G_A	**PDHX - AL356215.1**	22:21976934_T_C	**UBE2L3**
	7:128575797_TCTTAGCTATTGCTC_T	**NR**	8:11341870_G_T	**AF131216.5 - BLK**	2:191564757_G_T	**AC006460.1**
	17:43456728_C_T	**NR**	1:38276997_G_A	**MTF1**	8:11027441_G_A	**XKR6**
	GSA-rs11644034	**AC092723.5**	4:55548475_T_C	**KIT**	rs7579944	**H3P5 - LBH**
	rs1167796	**HIP1**	17:37921194_T_C	**IKZF3**	19:50121274_G_A	**PRR12**
	rs11697848	**KRT18P4 - RNU6-147P**	8:56980803_C_T	**RPS20**	1:173206297_G_A	**TNFSF4 - AL645568.3**
	16:68603798_C_T	**ZFP90**	rs3024505	**AL591846.1 - IL10**	6:138237989_C_T	**AL591468.1 - LINC02528**
	rs11724582	**IL2 - IL21**	rs3087243	**CTLA4 - ICOS**	5:133431834_C_T	**AC008608.1 - TCF7**
	rs11755393	**UHRF1BP1**	rs340630	**AFF1**	3:159694053_G_A	**IL12A-AS1**
	9:102337331_G_A	**AL162394.1, STX17-AS1, AL359710.1**	19:10463118_G_C	**TYK2**	rs7897633	**PRKG1**
	14:88383035_C_A	**GALC**	16:31272353_G_A	**ITGAM**	11:633689_T_C	**SCT - DRD4**
	rs11889341	**STAT4**	19:10477067_A_G	**TYK2**	11:128499000_T_C	**AP001122.1**
	GSA-rs12093154	**C1QTNF12**	6:34812701_T_G	**UHRF1BP1**	rs8023715	**RN7SKP181 - LINC02253**
	1:161492587_C_T	**AL590385.2, FCGR2A**	1:198594769_G_A	**AL157402.1 - PTPRC**	6:106569270_T_C	**ATG5**
	14:103263788_A_G	**TRAF3**	rs35677470	**DNASE1L3**	16:11038360_C_T	**CIITA - CLEC16A**
	1:114433946_G_A	**AP4B1-AS1**	16:85994484_C_T	**AC092723.5 - AC092723.4**	6:35033854_G_A	**ANKS1A**
	GSA-rs12531540	**JAZF1**	17:73341284_T_C	**AC011933.4, GRB2**	7:28185891_T_C	**JAZF1**
	4:965720_A_G	**DGKQ**	rs3734266	**UHRF1BP1**	rs887369	**TASL**
	4:8558266_C_T	**AC105345.1 - GPR78**	2:213871709_G_A	**IKZF2, AC093865.1**	rs911263	**RAD51B**
	8:71218360_G_A	**NCOA2**	11:71187679_C_T	**NADSYN1**	rs9270984	**HLA-DRB1 - HLA-DQA1**
	1:183532437_G_A	**SMG7, NCF2**	11:35098193_T_C	**PDHX - AL356215.1**	rs9271100	**HLA-DRB1 - HLA-DQA1**
	16:85966683_A_G	**AC092723.5**	rs4622329	**DRAM1**	6:32612301_T_C	**HLA-DQA1**
	19:18517767_G_A	**LRRC25 - SSBP4**	11:118573519_T_C	**AP002954.1**	3:58470351_C_T	**PDHB - AC116036.2**
	rs13385731	**RASGRP3**	rs4852324	**DGUOK-AS1**	rs9373839	**ATG5**
	rs1385374	**SLC15A4**	14:68731458_A_G	**RAD51B**	16:11174365_G_A	**CLEC16A**
	1:236039877_T_C	**LYST**				

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
AI thyroid	6:32798731_C_T	**TAP2**	8:129157016_A_G	**PVT1**	18:23598160_A_T	**SS18**
	10:124189197_A_G	**PLEKHA1**	9:21589041_G_A	**MIR31HG - H3P30**	17:40299024_AAGAAAGAAAGAC_A	**CTD-2132N18.3,RAB5C**
	6:408079_C_T	**IRF4**	7:100284913_C_G	**GIGYF1**	17:7240391_C_T	**ACAP1**
	11:308290_T_C	**IFITM2**	rs2234167	**TNFRSF14, TNFRSF14**	1:38377021_G_A	**INPP5B**
	8:128203024_T_TTC	**CASC19**	rs2251396	**AL645933.3 - MICA**	6:35536195_C_A	**AL033519.3 - FKBP5**
	19:18394549_T_TAAA	**NR**	12:56477694_A_T	**ERBB3**	18:12782849_C_G	**AP005482.2 - PTPN2**
	11:35317712_T_C	**SLC1A2**	rs2281388	**HLA-DPA2**	rs6535628	**AC002460.2**
	rs10748781	**AL391684.1 - LINC01475**	rs2294025	**TG**	6:90990050_G_A	**BACH2**
	rs1079418	**PDE10A**	rs229527	**C1QTNF6**	rs6832151	**AC095057.2 - LINC02265**
	12:9888642_T_A	**CLECL1 - CD69**	14:81458762_T_A	**TSHR**	5:76543768_G_A	**AC022414.1, PDE8B**
	12:111396249_A_G	**LINC01405**	1:157674997_G_T	**FCRL3 - AL356276.2**	6:148514301_G_T	**AL359382.2 - AL513164.1**
	17:45820723_C_T	**TBX21**	rs231775	**CTLA4**	10:6110875_C_A	**IL2RA - RPL32P23**
	3:108542960_G_A	**TRAT1**	8:128197088_A_G	**CASC19, PCAT1**	rs71198528	**HORMAD2**
	4:40307587_G_GCTTC	**NR**	5:133423495_T_C	**AC008608.1 - TCF7**	rs71508903	**ARID5B**
	5:110566360_C_T	**CAMK4**	rs2476601	**AL137856.1, PTPN22**	15:38850330_T_C	**RASGRP1**
	2:1407628_C_T	**TPO, AC105450.1**	rs2523989	**TRIM31**	17:35332740_T_A	**AATF, AATF**
	6:32389143_C_T	**TSBP1-AS1 - HLA-DRA**	rs2596449	**HCP5**	17:8841140_G_T	**PIK3R5**
	3:119221064_T_A	**TIMMDC1**	3:121825197_G_A	**CD86**	19:50162909_C_G	**IRF3**
	3:39320000_C_T	**CX3CR1**	20:17859706_A_G	**BANF2 - AL035045.1**	1:65429319_G_A	**JAK1**
	1:160419940_A_G	**VANGL2 - SLAMF6**	rs2757041	**RNASET2**	3:5015894_T_C	**BHLHE40-AS1**
	2:160552593_T_C	**AC009961.1**	21:16798586_T_A	**AJ009632.2**	4:26090375_A_G	**LINC02357**
	8:61518399_C_T	**RAB2A**	rs2857595	**NCR3 - UQCRHP1**	rs7537605	**VAV3**
	rs12117927	**EDARADD**	8:8308044_A_C	**AC103957.1 - AC103957.2**	2:191966452_G_C	**STAT4**
	1:167426424_G_A	**AL359962.3, CD247**	1:8497307_T_A	**RERE**	GSA-rs7594497	**PNPT1**
	14:98690394_A_T	**AL163932.1 - RN7SL714P**	rs312729	**CALM2P1 - AC011990.1**	12:9835925_G_T	**CLEC2D**
	12:103892941_G_T	**C12orf42 - AC084364.2**	rs3132613	**PRR3 - ABCF1**	rs763361	**CD226**
	rs1265883	**SLAMF6**	rs3184504	**ATXN2, SH2B3**	13:28604007_T_C	**FLT3**
	5:35837234_G_A	**AC137810.1 - IL7R**	10:64005582_T_TA	**RTKN2**	rs7679475	**QRFPR - TUBB4BP5**

	rs1270942	**CFB**	19:10463118_G_C	**TYK2**	6:112098896_T_C	**FYN**
	1:200840467_T_C	**RPL34P6 - GPR25**	rs35667974	**IFIH1**	rs7768644	**MUC22**
	rs12793348	**PANX1**	rs35677470	**DNASE1L3**	16:50335074_C_A	**ADCY7**
	3:188102104_G_T	**LPP**	rs3761959	**FCRL3**	1:108362169_GAC_G	**VAV3**
	8:134212652_C_T	**CCN4**	12:9833524_C_G	**CLEC2D**	16:8586245_T_C	**AC074052.2 - TMEM114**
	rs1352714	**DCHS2**	rs3775291	**TLR3**	rs886424	**LINC00243**
	rs1456988	**LINC01550 - AL163932.1**	rs3784099	**RAD51B**	9:5438435_C_A	**PLGRKT - CD274**
	1:113980577_T_C	**MAGI3**	rs3893464	**HLA-W - MICD**	rs925489	**PTCSC2**
	rs1521	**HLA-S - AL645933.3**	rs4248154	**MUC22**	1:156784982_T_C	**SH2D2A**
	2:12644736_C_T	**MIR3681HG**	4:10716939_G_C	**AC084048.1 - LINC02498**	6:31024601_C_A	**HCG22, HCG22**
	8:129569371_G_A	**LINC00824, LINC00824**	rs4313034	**ZNRD1ASP**	6:31091967_A_G	**PSORS1C1**
	rs1569723	**RPL13P2 - CD40**	1:25351581_G_A	**MIR4425 - LINC02793**	6:32605189_A_G	**HLA-DQA1**
	6:31335336_G_A	**AL671883.3**	rs4409785	**AP000820.2**	16:30930983_C_T	**FBXL19-AS1**
	13:24786915_T_G	**AL359736.1, SPATA13**	rs4713693	**MLN - LINC01016**	rs9355610	**RNASET2 - Z94721.1**
	rs1980493	**BTNL2, TSBP1-AS1**	rs4947296	**RNU6-1133P - C6orf15**	rs9394159	**ITPR3**
	rs1990760	**IFIH1**	exm-rs505922	**ABO, ABO**	6:136115507_T_C	**LINC00271**
	5:102622453_G_T	**MACIR - AC010406.1**	16:31185882_G_A	**AC009088.4 - NDUFA3P6**	13:43047260_G_A	**LINC02341**
	7:37409939_T_TGAAAAGG	**ELMO1**	3:105935232_TA_T	**NR**	4:40291916_A_G	**AC095057.2 - LINC02265**
	rs2069556	**TG**	rs5912838	**NR**	rs9878908	**GSTM5P1 - PPARG**

AD trait	SNP	Mapped gene	SNP	Mapped gene	SNP	Mapped gene
Vitiligo	6:32591291_T_A	**HLA-DRB1 - HLA-DQA1**	22:41767486_T_C	**TEF**	rs41342147	**FARP2**
	6:29937335_GTTTA_G	**NR**	1:8484823_A_G	**RERE-AS1, RERE**	13:43070206_T_G	**LINC02341 - TNFSF11**
	6:32440321_CTG_C	**HLA-DRB9**	rs4908760	**RERE**	rs71508903	**ARID5B**
	11:89017961_G_A	**TYR**	rs72928038	**BACH2**	20:49123043_A_G	**COX6CP2 - PTPN1**
	16:90026512_T_C	**DEF8**	19:4831878_A_G	**TICAM1 - AC027319.1**	9:127071493_T_C	**NEK6**
	rs229527	**C1QTNF6**	rs2456973	**IKZF4**	4:10726853_C_T	**AC084048.1 - LINC02498**

	3:188089254_C_T	**LPP**	15:28535266_A_C	**HERC2, HERC2**	rs3757247	**BACH2**
	21:43851828_T_A	**UBASH3A**	rs4409785	**AP000820.2**	3:196347253_T_A	**AC092933.2 - LINC01063**
	rs706779	**IL2RA**	8:133931055_A_T	**TG**	4:102223386_C_A	**PPP3CA**
	GSA-rs117744081	**CPVL**	10:115488331_G_A	**CASP7**	6:2908591_A_G	**SERPINB9 - AL133351.2**
	rs2111485	**FAP - IFIH1**	rs13136820	**AC095057.2 - LINC02265**	3:23512312_TA_T	**UBE2E2**
	GSA-rs11021232	**AP000820.2**	2:204740675_G_A	**CTLA4 - ICOS**	rs3814231	**CASP7**
	12:111833788_G_A	**LINC02356**	rs12203592	**IRF4**	11:128632555_T_C	**FLI1**
	rs3823355	**MICD - HCG9**	rs8083511	**TNFRSF11A**	16:89042861_A_G	**CBFA2T3**
	rs6059655	**RALY**	rs59374417	**CD80 - AC073352.2**	11:64021605_G_A	**PLCB3**
	2:55845109_T_C	**PPP4R3B, AC015982.2**	rs4822024	**ZC3H7B - TEF**	rs11966200	**SLC44A4**
	3:71557945_C_A	**FOXP1, AC097634.4**	rs11079035	**AC099811.2, RAB5C**	rs9468925	**HLA-B**
	rs2476601	**AL137856.1, PTPN22**	rs16843742	**PTPRC, PTPRC**	rs2236313	**RNASET2, AL159163.1**
	11:35274829_A_G	**SLC1A2**	rs10768122	**SLC1A2**	12:56401085_G_A	**SUOX - IKZF4**
	1:172715702_A_G	**FASLG - SLC25A38P1**	19:50168871_G_A	**BCL2L12, IRF3**	rs3213758	**RPGRIP1L**
	rs8192917	**AL136018.1, GZMB**	2:112010486_T_C	**MIR4435-2HG**	rs638893	**AP004609.4 - SETP16**
	rs1464510	**LPP**	rs1417210	**SLC29A3**		