
Genomic profiling of response to *in vivo* immune perturbations

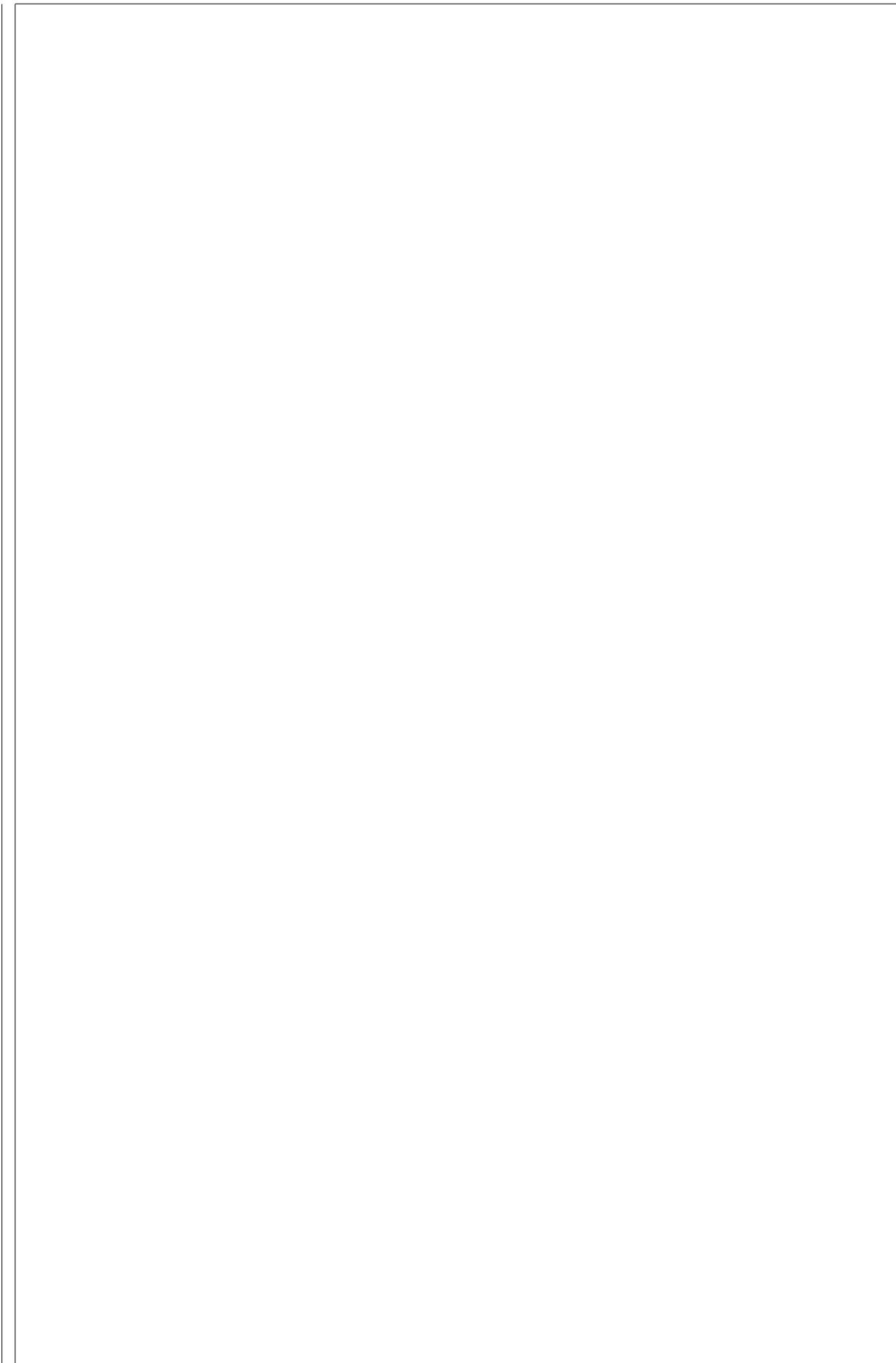
Benjamin Yu Hang Bai



Churchill College
University of Cambridge
Wellcome Sanger Institute

November 2020

This thesis is submitted for the degree of Doctor of Philosophy



Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit of 60 000 words for the Biology Degree Committee.

Abstract

The human immune system plays a central role in defense against infection, but its dysregulation is implicated in immune-mediated diseases. The past decade has seen increasing application of high-throughput technologies to profile, predict, and understand immune response to perturbation. The ability to measure immune gene expression at scale has led to the identification of transcriptomic signatures that predict clinical phenotypes such as antibody response to vaccines. It has also been recognised that both expression and phenotypic responses are traits with complex genetic architectures. This thesis examines the longitudinal transcriptomic response to immune perturbations, and its association with clinical response phenotypes and common genetic variation.

Chapter 2 explores transcriptomic response to pandemic influenza vaccine in a multi-ethnic cohort of healthy adults: the Human Immune Response Dynamics (HIRD) cohort. The success of vaccination in controlling influenza is indisputable, but it is incompletely understood why some individuals fail to mount protective antibody responses. I meta-analysed blood microarray and RNA sequencing (RNA-seq) datasets, identifying a distinct transition from innate immune response at day 1 after vaccination to adaptive immune response at day 7. Heterogeneity between measurement platforms made it difficult to identify single-gene transcriptomic associations with antibody response. Using a gene set approach, I found expression modules related to the inflammatory response, the cell cycle, CD4⁺ T cells, and plasma cells to be associated with vaccine-induced antibody response.

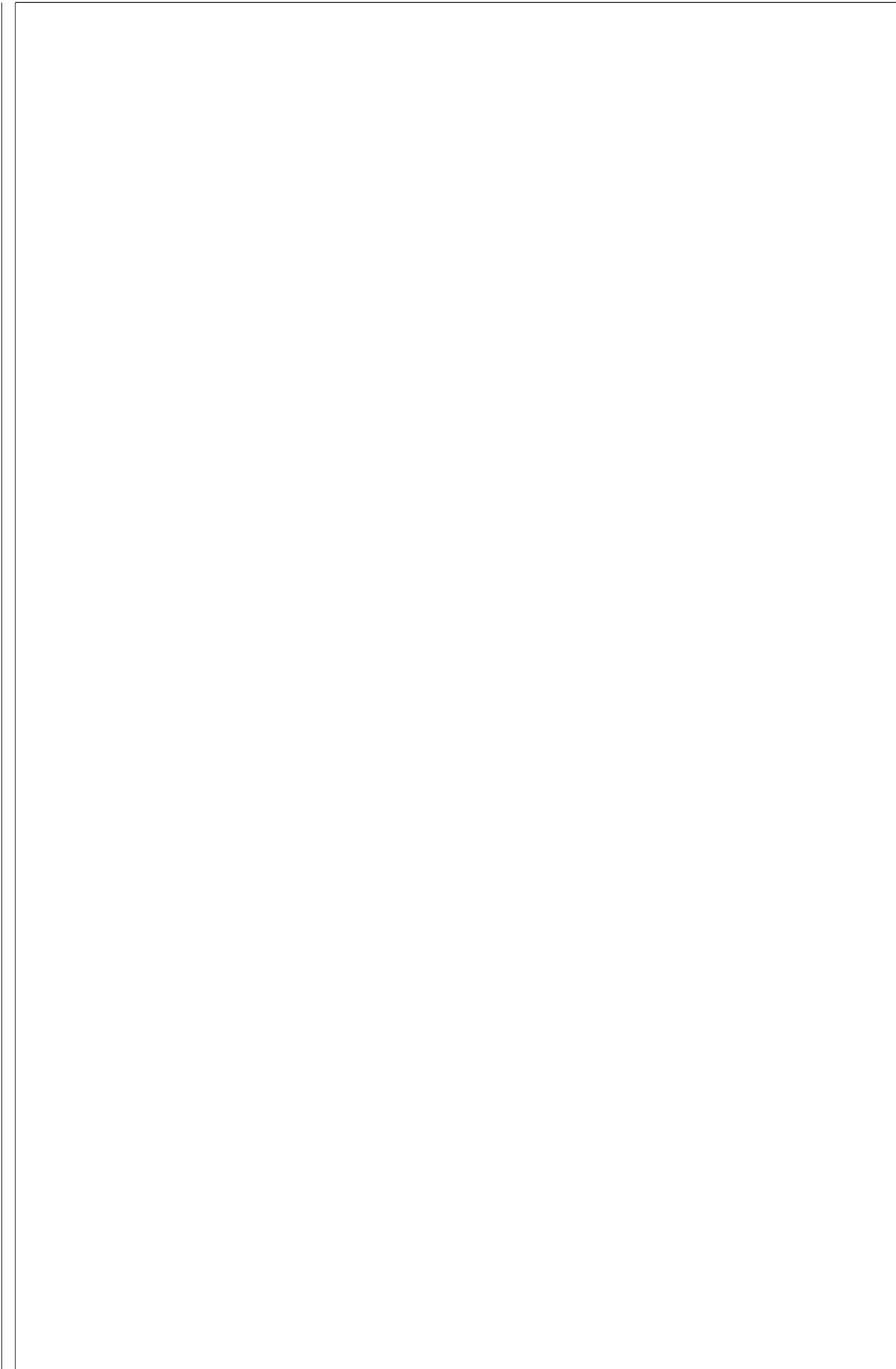
In Chapter 3, I map response expression quantitative trait loci (reQTLs) in the HIRD cohort to investigate regulation of transcriptomic response by common genetic variants. Rather than driving differential expression post-vaccination, the strongest reQTLs appeared to be explained by changes in cell composition revealing cell type-specific expression quantitative trait locus (eQTL) effects. For example, a reQTL identified for *ADCY3* specific to day 1 may be explained largely by high monocyte proportions at day 1 compared to other timepoints. Changes in cell composition present a significant challenge to interpreting reQTLs found through bulk sequencing of heterogeneous tissues.

Finally, Chapter 4 applies an analogous longitudinal study design to explore drug response in the Personalised Anti-TNF Therapy in Crohn's Disease (PANTS) cohort, a cohort of Crohn's disease (CD) patients treated with the anti-tumour necrosis factor (TNF) drugs, infliximab and adalimumab. Anti-TNF treatment has revolutionised patient care for CD, but 20–40 % of patients show primary non-response soon after starting treatment. I identified baseline expression modules associated with primary non-response, but also found significant heterogeneity of associations between the two drugs. Expression changes post-treatment in non-responders were

largely magnified in responders, suggesting there may be a continuum of response. Distinct expression trajectories identified for responders and non-responders revealed sustained expression differences during the first year of treatment. A set of interferon-related genes were regulated in opposing directions in responders and non-responders, presenting an attractive target for future studies of the biological mechanisms underlying non-response.

*A little learning is a dangerous thing;
Drink deep, or taste not the Pierian spring:
There shallow draughts intoxicate the brain,
And drinking largely sobers us again.*

Alexander Pope, *An Essay on Criticism*



Acknowledgements

“The more you know, the more you know you don’t know”. No other aphorism may be more apt at describing my perspective on these past four years, but what I do know—without a doubt—is that I owe a great deal to a great many people.

First and foremost, I must thank my supervisor, Carl Anderson. I greatly admire your commitment to embodying a culture of rigorous first-class science, and equally your commitment to the prosperity and well-being of everyone under your wing. I am indebted for all of your support and guidance, and your empathy and optimism during interesting times. Thank you for the privilege of the past four years—it has been an honour and a pleasure.

I must also take the opportunity to address the members of the Anderson team, past and present. The team has greatly evolved these past few years both in scientific direction and in membership, but it has always remained a friendly and enriching place to be. Whether over the lunch table in Murray’s, during our legendary team retreats, or in our traditionally lively team meetings, there was no lack of supportive and productive discussion. In particular, I must thank Aleksejs Sazonovs, Carla Jones-Bell, Elizabeth Goode, Laura Fachal, Leland Taylor, Loukas Moutsianas, Nikolaos Panousis, and Velislava Petrova, each of whom lent their ear to me countless times.

The research presented in this thesis would not have been possible without the help of numerous individuals and organisations. The projects in this thesis are highly collaborative, extending from the Wellcome Sanger Institute to London, Exeter, and the US. To Adrian Hayday, Nicholas Kennedy, Tariq Ahmad, and the team at AbbVie, thank you for granting me the opportunity to be part of these amazing collaborations. I owe a sincere gratitude to the many clinicians, scientists, and administrators who laid the foundation for this thesis, and even more so to the patients who contributed their samples. I would also like to thank the Wellcome Trust and other funding agencies that contributed to the projects in my thesis for their generous financial support.

Thank you to my thesis committee, Daniel Gaffney and Michael Inouye, who provided valuable feedback throughout the course of my PhD. I also extend my gratitude to the amazing research administrators I’ve worked with at Sanger: Carol Dunbar, Sally Bygraves, Eloise Stapleton, Paris Litterick, Rachel Henry, and Sophie Leggett. I’m sure there has been many a time you’ve known more about my projects than me. To the Sanger Sample Management, Pipelines, and Informatics teams, the incredible work you do to keep the sequencing and computing resources running smoothly was critical for this thesis, and for all science done on the Campus.

On the other side of the balance, I would like thank all the friends I have made during my time in Cambridge. To my fellow PhD cohort, I’ll remember all the punting trips, the pub quizzes,

the eSCAMPS dinners, and our shared commiserations on the Campus bus. To the Churchill College badminton team, thank you for having me, and best wishes for all the training sessions and league matches in the years to come. I would like to give a special mention to my friends from the Cambridge University Anime and Manga Society. Many good times were had; many good memories were made. You've truly made my "campus life" very much multicoloured.

Finally, to my family—to my brother, my mother, my father, my grandparents—I could always depend on your unconditional love and support. Despite the thousands of miles between us, all of you were held close to my heart.

Contents

Declaration	iii
Abstract	v
Acknowledgements	ix
Contents	xi
List of Figures	xv
List of Tables	xix
1 Introduction	1
1.1 Genetic association studies for complex traits	1
1.1.1 Structure and variation of the human genome	1
1.1.2 Lessons from the past fifteen years	2
1.1.3 From complex trait to locus	3
1.1.4 From locus to causal variant	5
1.1.5 From causal variant to target gene	5
1.2 Gene expression as an intermediate molecular phenotype	5
1.2.1 Regulation of gene expression	5
1.2.2 Expression quantitative trait loci (eQTLs)	6
1.2.3 Context-dependent eQTLs	7
1.2.4 Response eQTLs (reQTLs)	7
1.2.5 Gene prioritisation using eQTLs	10
1.3 Phenotypes of immune response	10
1.3.1 An overview of the immune system	10
1.3.2 High-throughput immunology	11
1.4 Thesis outline	12
2 Transcriptomic response to Pandemrix vaccine	15
2.1 Introduction	15
2.1.1 Influenza	15
2.1.2 Seasonal influenza vaccines	16
2.1.3 Quantifying immune response to influenza vaccines	17

2.1.4	Systems vaccinology of seasonal influenza vaccines	18
2.1.5	The Human Immune Response Dynamics (HIRD) cohort	20
2.1.6	Chapter summary	21
2.2	Methods	22
2.2.1	Existing HIRD data and additional data generation	22
2.2.2	Computing baseline-adjusted measures of antibody response	22
2.2.3	Genotype data generation	25
2.2.4	Genotype data preprocessing	25
2.2.5	Computing genotype principal components as covariates for ancestry	25
2.2.6	RNA-seq data generation	29
2.2.7	RNA-seq quantification and preprocessing	29
2.2.8	Array data preprocessing	32
2.2.9	Differential gene expression (DGE)	35
2.2.9.1	Platform and batch effects	35
2.2.9.2	Per-platform DGE model	38
2.2.9.3	Choice of DGE meta-analysis method	38
2.2.9.4	Prior for between-study heterogeneity	40
2.2.9.5	Prior for effect size	40
2.2.9.6	Example of priors	40
2.2.9.7	Multiple testing correction	41
2.2.10	Ranked gene set enrichment using blood transcription modules	41
2.3	Results	43
2.3.1	Extensive global changes in expression after vaccination	43
2.3.1.1	Innate immune response at day 1 post-vaccination	43
2.3.1.2	Adaptive immune response at day 7 post-vaccination	45
2.3.2	Expression associations with antibody response	45
2.3.2.1	Between-platform heterogeneity hinders detection of gene-level associations	45
2.3.2.2	Module-level associations with antibody response	47
2.4	Discussion	52
3	Genetic architecture of transcriptomic response to Pandemrix vaccine	55
3.1	Introduction	55
3.1.1	Host genetic factors affecting influenza vaccine response	55
3.1.2	reQTLs induced by influenza vaccination	56
3.1.3	Chapter summary	57
3.2	Methods	58
3.2.1	Overall strategy for reQTL mapping	58
3.2.1.1	Adjusting for population structure using linear mixed models	58
3.2.1.2	Multi-condition models	59
3.2.1.3	Additional expression preprocessing	60
3.2.2	Genotype phasing and imputation	61
3.2.3	Estimation of kinship matrices	63

3.2.4	Estimation of cell type abundance from expression	63
3.2.5	Finding unmeasured covariates using factor analysis	67
3.2.6	eQTL mapping per timepoint	70
3.2.7	Joint eQTL analysis across timepoints	70
3.2.8	Defining shared eQTLs and reQTLs	71
3.2.9	Replication of eQTLs in a reference dataset	72
3.2.10	Genotype interactions with cell type abundance	73
3.2.11	Gene set enrichment analyses	75
3.2.12	Statistical colocalisation	75
3.3	Results	76
3.3.1	Mapping reQTLs in the HIRD cohort	76
3.3.2	Characterising reQTLs post-vaccination	77
3.3.3	Exploring possible mechanisms generating reQTLs	81
3.3.3.1	Differential expression of genes with reQTLs	81
3.3.3.2	Genotype by cell type abundance interaction effects	81
3.3.3.3	Colocalisation with external QTL datasets at the <i>ADCY3</i> locus	83
3.4	Discussion	91
4	Transcriptomic associations with anti-TNF drug response in Crohn's disease patients	95
4.1	Introduction	95
4.1.1	Crohn's disease and inflammatory bowel disease	95
4.1.2	Anti-TNF therapies for Crohn's disease	96
4.1.3	Anti-TNF treatment failure	97
4.1.4	Predicting patient response to anti-TNFs	97
4.1.5	Chapter summary	98
4.2	Methods	99
4.2.1	The Personalised Anti-TNF Therapy in Crohn's Disease (PANTS) cohort	99
4.2.2	Definition of timepoints	99
4.2.3	Definition of primary response and primary non-response	100
4.2.4	Library preparation and RNA-seq	101
4.2.5	RNA-seq quantification and preprocessing	102
4.2.6	Differential gene expression	102
4.2.6.1	Variable selection by variance components analysis	102
4.2.6.2	Contrasts for pairwise group comparisons	106
4.2.6.3	Spline model of expression over time	107
4.2.6.4	Clustering expression over all timepoints	108
4.2.6.5	Gene set enrichment analyses	108
4.2.7	Genotyping and genotype data preprocessing	108
4.2.8	reQTL mapping	109
4.2.8.1	Computing genotype principal components	109
4.2.8.2	Finding hidden confounders in expression data	109
4.2.8.3	Computing kinship matrices	109

4.2.8.4	Mapping eQTLs per timepoint	111
4.2.8.5	Joint reQTL mapping over all timepoints	111
4.3	Results	112
4.3.1	Longitudinal RNA-seq data from the PANTS cohort	112
4.3.2	Baseline gene expression associated with primary response	112
4.3.3	Assessing previously reported baseline predictors of primary response	114
4.3.4	Post-induction gene expression associated with primary response	118
4.3.5	Magnification of expression changes from baseline to post-induction in responders	119
4.3.6	Interferon modules with opposing expression changes in responders and non-responders	119
4.3.7	Sustained expression differences between primary responders and non-responders during maintenance	123
4.3.8	Limited evidence for changes in genetic architecture of gene expression over time	127
4.4	Discussion	133
5	Discussion	141
5.1	Strategies for detecting robust associations	141
5.2	Responder analysis	143
5.3	Challenges in the interpretation of bulk expression data	144
5.4	From association to prediction	146
5.5	From association to causality	147
5.6	Triangulation	149
5.7	Concluding remarks	149
Bibliography		151
List of Abbreviations		185

List of Figures

1.1	The mosaic structure of human genetic variation.	3
1.2	Effect size and allele frequency of trait-associated genetic variants.	4
1.3	Types of tissue-dependent <i>cis</i> -expression quantitative trait locus (eQTL) effects.	8
1.4	High-throughput technologies for systems immunology.	12
2.1	Overview of Human Immune Response Dynamics (HIRD) study data.	23
2.2	Antibody titre data and responder definitions.	24
2.3	Distribution of patient titre response indexs (TRIs), stratified by expression measurement platform.	27
2.4	Sample filters for marker missingness and marker heterozygosity rate.	28
2.5	HIRD samples projected onto principal component (PC) axes defined by principal component analysis (PCA) of HapMap 3 samples.	30
2.6	FastQC per-base sequence quality (Phred scores) versus read position for RNA sequencing (RNA-seq) samples.	31
2.7	FastQC per-read GC distributions for RNA-seq samples.	31
2.8	Distributions of removed short non-coding RNA (ncRNA) and globin counts as a proportion of total counts in RNA-seq samples.	33
2.9	Distribution of the proportion of samples in which genes were detected (non-zero expression).	33
2.10	Distributions of gene expression for RNA-seq samples before and after filtering low expression and non-detected genes.	34
2.11	Distribution of raw foreground intensities for HIRD array samples ($n = 173$).	34
2.12	Distribution of per-sample expression estimates after normalisation and collapsing of probes to genes.	36
2.13	First four standardised PCs in the expression data, colored by array batch/RNA-seq pool (a, c), timepoint (b, d), and binary response status (e, f).	37
2.14	Gamma prior for τ (blue) used for <code>bayesmeta</code> analyses of the day 1 versus baseline effect, compared to the empirical distribution of per-gene frequentist <code>metafor::rma.uni</code> estimates for τ .	42
2.15	Normal prior for μ (blue) used for <code>bayesmeta</code> analyses of the day 1 versus baseline differential gene expression (DGE) effect, compared to the empirical distribution of per-gene frequentist <code>metafor::rma.uni</code> estimates for τ .	42

2.16 Normalised gene expression for 857 genes differentially expressed between any pair of timepoints ($\text{lfsr} < 0.05$, $ \text{FC} > 1.5$).	44
2.17 Transcriptomic modules up or downregulated between pairs of timepoints.	46
2.18 DGE effect sizes ($\log_2 \text{FC}$) estimated in array versus RNA-seq samples, colored by significance in frequentist random effects meta-analysis using <code>rma.uni</code> at BH FDR < 0.05	48
2.19 DGE effect sizes ($\log_2 \text{FC}$) estimated in array versus RNA-seq samples, colored by significance in Bayesian random effects meta-analysis using <code>bayesmeta</code> at <code>ashr</code> LFSR < 0.05	49
2.20 Estimates of between-platform heterogeneity τ from frequentist and Bayesian meta-analysis, for the 58 genes with a significant association between day 7 expression and binary responder/non-responder status in Sobolev <i>et al.</i> [162].	50
2.21 Gene expression modules associated with antibody response (TRI).	51
3.1 Simulating the effect of data transformation on response expression quantitative trait locus (reQTL) effects.	62
3.2 Correlation matrix of standardised <code>xCell</code> cell type enrichment scores in HIRD array and RNA-seq datasets.	65
3.3 Contribution of each cell type score to each PC dimension after PCA of standardised <code>xCell</code> cell type enrichment scores.	66
3.4 Comparison of standardised <code>xCell</code> scores with normalised HIRD fluorescence-activated cell sorting (FACS) measurements, for monocytes, natural killer (NK) cells, and plasma cells.	68
3.5 Correlation of known variables to the first 25 PEER factors estimated from the array and RNA-seq mega-analysis expression data at baseline.	69
3.6 Number of significant genes with an eQTL detected on chromosome 1 as a function of the number of PEER factors included as covariates.	71
3.7 Replication rate π_1 of HIRD eQTLs in GTEx whole blood eQTL reference data.	74
3.8 Summary of HIRD eQTL mapping from mega-analysis of array and RNA-seq expression data, binned by patterns of lead variant significance over the three timepoints.	78
3.9 eGenes where the lead eQTL was a reQTL between a pair of timepoints.	79
3.10 z -statistic for difference in beta post-vaccination versus baseline for shared and reQTLs, against distance from the eGene transcription start site (TSS).	80
3.11 Expression and lead eQTL of <i>ADCY3</i> over study timepoints.	82
3.12 Expression and lead eQTL of <i>SH2D4A</i> over study timepoints.	83
3.13 Effect of estimated monocyte abundance on <i>ADCY3</i> expression at baseline, stratified by genotype at a day 1 <i>ADCY3</i> reQTL.	84
3.14 Effect of estimated monocyte abundance on <i>ADCY3</i> expression at day 1, stratified by genotype at a day 1 <i>ADCY3</i> reQTL.	85
3.15 Expression of <i>ADCY3</i> in sorted immune cell subsets.	87
3.16 Sensitivity analysis for multi-trait colocalisation at the <i>ADCY3</i> locus.	89
3.17 Multi-trait colocalisation at the <i>ADCY3</i> locus.	90

4.1	Sample size and study day distribution for Personalised Anti-TNF Therapy in Crohn's Disease (PANTS) study RNA-seq samples, stratified by timepoint and study group.	101
4.2	Distribution of RNA-seq samples from each patient among timepoints.	103
4.3	Correlation matrix of variables measured in PANTS that were considered as potential predictor variables.	104
4.4	Variance components analysis showing the distribution of per-gene percentage of variance in expression explained by each variable.	106
4.5	1000 Genomes Project (1000G) samples and PANTS samples projected onto 1000G genotype PC1 and PC2 axes, colored by (a) superpopulation and (b) population. .	110
4.6	Number of eGenes on chromosome 1 vs. number of PEER factors included in eQTL mapping as covariates.	111
4.7	Volcano plots of DGE between primary responders (PR) and non-responders at week 0; unadjusted (top row) and adjusted (bottom row) for cell composition; for infliximab (IFX), adalimumab (ADA), or with both drugs pooled.	115
4.8	Top modules differentially expressed between primary responders (PR) and non-responders at week 0, unadjusted for cell composition.	116
4.9	Top modules differentially expressed between primary responders (PR) and non-responders at week 0, adjusted for cell composition.	117
4.10	Volcano plots of DGE between primary responders (PR) and non-responders at week 14; unadjusted (top row) and adjusted (bottom row) for cell composition; for infliximab (IFX), adalimumab (ADA), or with both drugs pooled.	120
4.11	Top modules differentially expressed between primary responders (PR) and non-responders at week 14, unadjusted for cell composition.	121
4.12	Top modules differentially expressed between primary responders (PR) and non-responders at week 14, adjusted for cell composition.	122
4.13	Expression log ₂ FC from week 0 to week 14 in primary responders (PR) versus non-responders (PNR), for genes that differentially expressed from week 0 to week 14 in both responders and non-responders, with a significantly different effect size between responders and non-responders.	123
4.14	Top modules differentially expressed between week 14 and week 0, unadjusted for cell composition.	124
4.15	Top modules differentially expressed between week 14 and week 0, adjusted for cell composition.	125
4.16	tmod evidence plots showing interferon-related modules specifically upregulated from week 0 to week 14 in primary non-responders (PNR), but not in primary responders (PR).	126
4.17	Gap statistic versus cluster number <i>k</i>	128
4.18	Normalised expression over the timepoints for genes in the six identified clusters. .	129

4.19 Overlap of genes differentially expressed between responders and non-responders from the spline model, the week 14 responder versus non-responder contrast (w14), and the interaction between week 0 to week 14 change and response status contrast (w14 – w0).	130
4.20 Post-treatment eQTL effect sizes at week 30 and week 54 compared to baseline effect sizes at week 0.	132
5.1 The three assumptions of Mendelian randomisation (MR).	149

List of Tables

2.1	Descriptive statistics for Human Immune Response Dynamics (HIRD) individuals with both expression and antibody data.	26
2.2	Distribution of HIRD samples among timepoint and responder groups in the array batches and RNA sequencing (RNA-seq) pools.	39
4.1	Patient characteristics for the Personalised Anti-TNF Therapy in Crohn's Disease (PANTS) RNA-seq subcohort.	113

--

Chapter 1

Introduction

Observable human characteristics or traits are called phenotypes. Variation in phenotype emerges from the interplay of genetics, environment, and pure chance. The contributions of each can vary immensely from phenotype to phenotype. Traits for which genetic variation explains a non-zero percentage of phenotypic variation are heritable. Virtually all phenotypic traits are heritable to some degree, and twin studies provide upper bounds on this heritability by partitioning phenotypic variation into genetic and environmental components [1].

Genetic variation presents a unique opportunity to probe the causal molecular mechanisms underlying phenotypes. Information encoded in the genome has phenotypic consequences only after flowing through multiple molecular layers. This guiding principle is the central dogma of molecular biology, whereby the flow is directed from DNA to RNA to protein via transcription and translation. Barring somatic mutation, an individual's genome is fixed at conception, providing a causally upstream anchor that can be measured with relatively little error. A mainstay of the field of human genetics is uncovering the specific genetic variants that contribute to heritability of phenotypes through statistical association of variants and phenotypes. Although not immune to population-level biases like stratification [2] and collider bias [3], genetic association has intrinsic resistance to reverse causality, an issue that permeates observational studies of the causes of human phenotypes.

1.1 Genetic association studies for complex traits

1.1.1 Structure and variation of the human genome

The human genome is over three billion bp (base pairs) in length, containing 20 000–25 000 protein-coding genes that span 1–3 % of its length, with the remaining sequence being non-coding [4, 5]. Each diploid cell contains two copies of the genome, organised into 46 chromosomes comprised of 23 maternal-parental pairs: 22 pairs of homologous autosomes and one pair of sex chromosomes. Variation in the genome between individuals in a population exists in the form of single nucleotide polymorphisms (SNPs), short indels, and structural variants. For common population variants with minor allele frequency (MAF) >1 %, the vast majority (>99.9 %) are SNPs and short indels [5]. On average, a pair of genomes differs by one SNP per 1000–2000 bp [6]. Each version of a variant is called an allele; each individual has a maternal and parental

allele at each variant.

The large number of variants in a population are inherited on a smaller number of haplotypes: contiguous stretches of the genome passed through generations via meiotic segregation. The fundamental sources of genetic diversity, mutation and meiotic recombination, generate new alleles and break apart haplotypes into shorter ones over evolutionary time. Variants that are physically close on a chromosome are less likely to flank a recombination event, hence are more likely to cosegregate from parent to offspring on the same haplotype (genetic linkage). Genetic linkage is one source of **linkage disequilibrium (LD)**: the non-random association of alleles at two variants, differing from expectation based on their population frequencies and the law of independent assortment. LD can be quantified by r^2 , the squared correlation coefficient between alleles in a specific population [7]. Recombination events are not distributed uniformly throughout the genome. The genome is a mosaic of haplotype blocks delimited by recombination hotspots, characterised by strong LD within blocks, and little LD between blocks [8, 9] (Fig. 1.1). The structure of correlated haplotypes reflects a population's unique evolutionary history, and can be used to trace the demography of populations back through time [10].

1.1.2 Lessons from the past fifteen years

Genetic variants can affect heritable traits by impacting the function or regulation of target genes. How genetic variation contributes to a particular trait defines its genetic architecture: the number of genes affecting the trait; and the frequencies, effect sizes, and interactions of trait-associated alleles [12, 13]. The number of genes defines a spectrum of traits from monogenic (where inheritance follows simple Mendelian patterns) to polygenic (where inheritance is complex). Proposed architectures differ strikingly among complex traits, even for traits with phenotypic similarities like **type 1 diabetes (T1D)** and **type 2 diabetes (T2D)** [12]. Consistently, however, the number of genes and genetic variants affecting a complex trait is large (ranging from dozens to many thousands), the average effect size of trait-associated variants is small, and the contribution of environment is substantial [14–16].

Since the 1980s, linkage analysis has been used to map the chromosomal positions and regions (loci) affecting traits by tracing the cosegregation of markers (variants with known positions) with the trait in family pedigrees [17–19]. Linkage analysis was complemented by early genetic association studies, which largely focused on variants in or near candidate genes selected on the basis of prior biological knowledge [20]. These methods saw much success for Mendelian traits, but application to most complex traits proved challenging. Small average effect sizes meant penetrance was too low to reliably observe cosegregation in pedigrees [19]. Early candidate gene studies were severely underpowered to detect such small effects [21].

The past fifteen years have seen the rise of **genome-wide association studies (GWASs)** that systematically test common variants selected in a hypothesis-free manner across the whole genome (Fig. 1.2). Using large sample sizes to overcome small effects and the large multiple testing burden, thousands of associations have been discovered for complex traits and diseases, many robustly replicated across populations [19, 22]. A number of take-home messages have emerged. Most genetic variance is additive; the contributions of dominance and epistatic interaction are small [13]. Variants with effects on multiple phenotypes (pleiotropy) are widespread [19]. Even traits

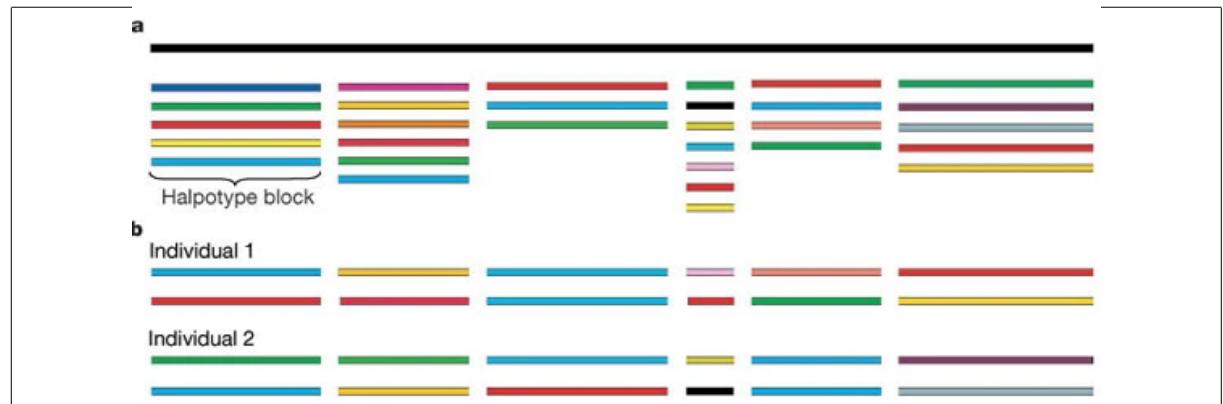


Figure 1.1: The mosaic structure of human genetic variation. Large parts of the genome can be divided into haplotype blocks between 5–200 kbp in length, with strong intra-block LD. For each block, three to seven common haplotypes (indicated by different colors) represent the majority of variation found in humans. An individual carries two haplotypes per block, one inherited from each parent. The exact structure and diversity of haplotype blocks varies between populations. Information on the haplotypes, their locations in the genome, and their frequencies in different populations form a “haplotype map” of the genome. Figure reprinted by permission from Springer Nature: Springer Nature, *Nature*, Pääbo [11], © 2003.

that are molecular rather than whole-organism phenotypes can be remarkably polygenic, with hundreds to thousands of associated loci [23]. GWAS sample sizes in the millions are increasingly commonplace, and the discovery of new associations with ever smaller effects as sample sizes increase shows no sign of plateauing [24, 25].

1.1.3 From complex trait to locus

GWASs rely on the tendency of common variants on the same haplotype to be in strong LD. As the number of haplotypes is relatively few, it is possible to select a subset of tag variants such that all other known common variants are within a certain LD threshold of that subset. In practice, there is enough redundancy that the number of variants measured on a modern genotyping array (in the order of 10^5 – 10^6) is sufficient to tag almost all common variants [27, 28]. Associations with unmeasured variants are indirectly detected through correlation with a tag variant. Furthermore, as unrelated individuals still share short ancestral haplotypes, study samples can be assigned haplotypes from a panel of haplotypes derived from reference samples by matching on directly genotyped variants. Genotypes at untyped variants can then be assigned from those haplotypes. This process—genotype imputation—allows ascertainment of many more variants than are directly genotyped [29] and helps to recover rarer variants that are poorly-tagged [22]. Modern imputation panels enable cost-effective GWASs testing tens of millions of variants as rare as 0.01–0.1 % in diverse populations [30].

Testing large numbers of variants incurs a massive multiple testing burden, but acknowledging the correlation between variants due to LD and restricting tests to common variants, there are only the equivalent of $\sim 10^6$ independent tests in the European genome, regardless of the number of tests actually performed [31]. The field has thus converged on a fixed discovery threshold of $0.05/10^6 = 5 \times 10^{-8}$ for genome-wide significance in European populations [32], akin to controlling the family-wise error rate (FWER) to below $\alpha = 0.05$ using the Bonferroni

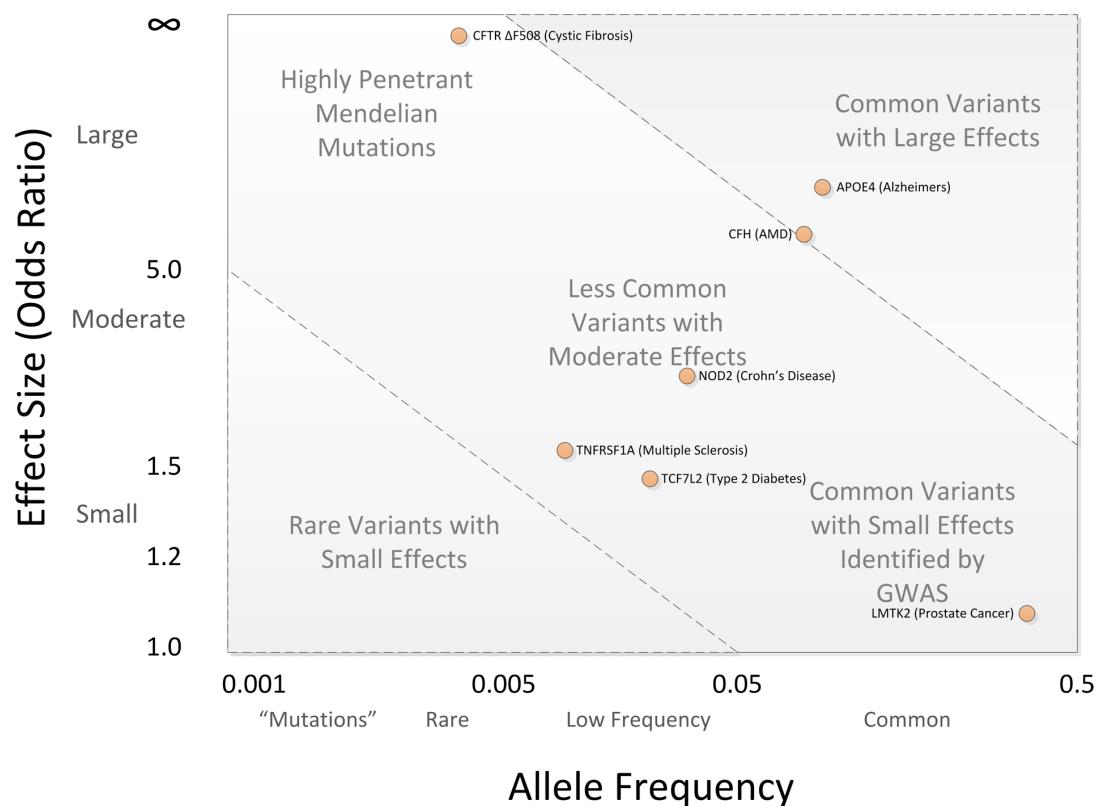


Figure 1.2: Effect size and allele frequency of trait-associated genetic variants. Different classes of genetic effects require different and complementary methods. Linkage analysis is suited to detecting Mendelian variants with large effects. GWAS is suited to detecting common variants with small effects. There are few common variants with large effects due to selection pressure. Rare variants with small effects are hard to distinguish from noise without very large samples. They are also poorly tagged by genotyping arrays and difficult to impute. Studies focusing on rare variants often employ whole-exome sequencing (WES) or whole-genome sequencing (WGS) to directly type variants. Figure reproduced from Bush *et al.* [26] under the CC BY 4.0 license (creativecommons.org/licenses/by/4.0/legalcode).

correction*.

1.1.4 From locus to causal variant

By design, a significant trait-associated variant from a **GWAS** needs not be a variant that causally affects the trait and may only tag a causal variant. The resolution of the associated locus depends on the local **LD** structure. Fine-mapping is the process of determining which of the many correlated variants in an associated locus are most likely to be causal, assuming the causal variants are observed either by direct genotyping or confident imputation. Due to incomplete power, the causal variants in a locus are not necessarily the ones with the strongest associations [34]. Bayesian fine-mapping methods take a variable selection approach, assigning each variant a posterior probability of causality. A credible set of variants likely to contain the causal variant in the locus with some probability can then be determined [34, 35]. The ability to separate causal and tag variants depends on factors including **LD**, sample size, and the effect size and number of causal variants [22, 34].

1.1.5 From causal variant to target gene

Most causal variants for Mendelian traits are coding variants (nonsense, missense, or frameshift) that impact protein sequence [36]. In contrast, over 90 % of **GWAS** loci fall in non-coding regions [37], and often too far from the nearest coding region to be in **LD** [38]. Even if the causal variants in a locus are fine-mapped, one of the greatest challenges following a **GWAS** is prioritising the target genes through which those variants affect the trait. A reasonable heuristic is to assign the gene with the nearest **transcription start site (TSS)** or body as the target, particularly for metabolite traits [39]. For improved accuracy across a variety of complex traits, integrative methods for gene prioritisation combine variant-to-gene distance with other metrics and data types drawn from numerous external sources [39–41].

1.2 Gene expression as an intermediate molecular phenotype

1.2.1 Regulation of gene expression

Gene regulation data are indispensable for gene prioritisation. Rather than directly impacting the coding sequence of a gene, many non-coding **GWAS** loci are hypothesised to affect traits by affecting the regulation of target gene expression [37, 42]. Unlike genotype, expression is dynamic across time and space. Diverse expression programs are responsible for the myriad of cell and tissue types generated during development, and enable adaptation in response to environmental stimuli.

Expression is the product of eukaryotic transcription, a multi-step process involving interactions between DNA, RNA, and hundreds of proteins [43]. Transcription of the **pre-messenger RNA (mRNA)** is initiated when RNA polymerase and **transcription factors (TFs)** form part

*The Bonferroni correction makes no assumptions about the dependence structure of the *p*-values, controlling the **FWER** (probability of at least one type I error) exactly under any structure. It is conservative (i.e. controls the **FWER** at a stricter level than the chosen threshold α) even for independent tests. In fact, it is always conservative unless the *p*-values have strong negative correlations [33].

of a protein complex around the promoter region and **TSS** of a gene. **TFs** can also bind to more distant *cis*-regulatory elements such as enhancers and repressors. These distant regulatory elements interact with the promoter region via DNA looping. Transcription can only happen in regions of open chromatin, where the packing of DNA-histone complexes (nucleosomes) is loose enough that the DNA is physically accessible to the transcriptional machinery. Chromatin accessibility is partially determined by histone modifications such as methylation, acetylation, phosphorylation, and ubiquitination [44]. The DNA itself can also be modified; methylation at CpG sites in promoters tends to repress transcription [45].

To form a mature **mRNA**, the pre-**mRNA** is capped at the 5' end by a modified nucleotide and at the 3' end by a poly(A) tail. Exons are joined by spliceosomes that cut and rejoin the pre-**mRNA** at one or more pairs of splice sites, excising the intronic sequence between each pair. The choice of splice sites determines which of many alternatively-spliced transcripts is produced. Post-transcriptional regulation of mature **mRNAs** is also possible via RNA editing [46] and regulatory elements in the flanking 5' and 3' untranslated regions (**UTRs**) [47].

In line with the regulatory hypothesis, **GWAS** variants are heavily enriched in regulatory elements annotated by functional genomics projects (e.g. ENCODE [4]), including regions of open chromatin, histone binding sites, **TF** binding sites, enhancers, splice sites, and **UTRs** [48–52]. Furthermore, this enrichment is often observed in particular contexts (tissues, cell types, or cell states [22, 37, 42]). An example is the enrichment of fine-mapped **SNPs** associated with risk of **immune-mediated inflammatory disease (IMID)** in CD4⁺ T cell enhancers, particularly in enhancers activated after stimulation [50]. These results put forth expression as an important intermediate that links non-coding **GWAS** variants to their associated traits, and helps nominate trait-relevant contexts and target genes.

1.2.2 Expression quantitative trait loci (eQTLs)

Expression is a complex molecular phenotype in itself, with a heritability of 15–30 % [53]. Genome-wide assays for expression, such as microarrays and **RNA sequencing (RNA-seq)**, were among the earliest high-throughput technologies developed for quantifying molecular phenotypes. Genetic loci associated with quantified gene expression are called **expression quantitative trait loci (eQTLs)**. Large-scale efforts such as the Genotype-Tissue Expression (GTEx) project [54] have pioneered the study of eQTLs and other **molecular quantitative trait loci (molQTLs)** over the past decade [55].

eQTL effect sizes are large relative to variants associated with whole-organism phenotypes, with the average eQTL explaining 5–18 % of additive genetic variance for its associated gene [53]. The eQTLs with the largest effects tend to be concentrated near the **TSS** of their target gene (*cis*-eQTLs), affecting **TF** binding sites and other local regulatory elements. eQTLs further away or on a different chromosome are called *trans*-eQTLs. The exact threshold separating *cis* from *trans* on the same chromosome is arbitrary; <1 Mbp and >5 Mbp are commonly used thresholds for *cis*- and *trans*-eQTLs respectively [56–58]*. In general, eQTL effect size declines with distance

*Having a threshold is often a matter of practicality to reduce the number of variants tested. Assaying expression is still more costly than array genotyping, so eQTL mapping sample sizes are small compared to **GWASes**. Even though eQTL effects are relatively large, eQTL mapping genome-wide would be equivalent to performing **GWASes** on thousands of continuous phenotypes, incurring enormous computational and multiple testing burdens.

to the TSS, and *trans*-eQTLs have smaller effects compared to *cis*-eQTLs [55]. *Trans*-eQTLs often represent *cis*-eQTLs of regulatory molecules like TFs and RNA-binding proteins that may target many genes in *trans* as master regulators [57, 59]. Gathering large enough samples to detect *trans*-eQTLs remains a priority, as most expression heritability is driven by *trans* rather than *cis* effects, perhaps due to small but wide-reaching effects [60].

1.2.3 Context-dependent eQTLs

Like expression itself, the effects of eQTLs are highly context-dependent [55, 57]. When the effect size of an eQTL is not the same in all environments, but differs depending on the environment, the eQTL is said to interact with those environments. This can manifest as a statistical interaction in a regression model with a multiplicative genotype-environment term, where the effects of environment and genotype on expression are not additive at the chosen scale for measuring expression. A non-exhaustive list of environmental contexts that have been found to interact with eQTLs includes sex [61], age [61], ancestry [62–64], tissue [65, 66], purified cell type [62, 67–70], cell type composition in bulk samples [71–74], cell differentiation stage [75], disease status [68], and experimental stimulation (see Section 1.2.4). These contexts can be interdependent; for example, tissue-dependent effects may arise from a combination of cell type-dependence and varying cell composition between tissues.

A multitude of molecular mechanisms could facilitate genotype-environment interactions at eQTLs. Fu *et al.* [76] mapped eQTLs in blood and four non-blood tissues (Fig. 1.3), and proposed mechanisms that might explain discordant effects of an eQTL allele on a target gene between tissues, assuming the eQTL disrupts a regulatory factor's binding site. Different effect sizes of same or opposite signs could arise from tissue-dependent effects of the same factor, such as activating expression in one tissue and repressing it in another (e.g. due to cofactors, or from binding of different factors in different tissues at the same site). Effects specific to a tissue could arise from tissue-specific expression of a regulatory factor. A tissue-specific effect could also reflect tissue-specific target gene expression, as the eQTL effect will be zero in a tissue where the target is not expressed (e.g. due to chromatin inaccessibility). Tagging of different causal variants in the two tissues, potentially with differing tagging efficiency (i.e. LD), could also generate the above scenarios [76]. Furthermore, the complexity of human gene regulation means these mechanisms might be acting at epigenetic, pre-, co-, or post-transcriptional regulatory levels [53]. Detection of context-dependent effects merely exposes differences in regulatory architecture between contexts. Much like in GWASs, going from association to underlying mechanism requires considering data types beyond just genotype and expression.

1.2.4 Response eQTLs (reQTLs)

A important class of context-dependent eQTLs are response expression quantitative trait loci (reQTLs), where the interacting environment is experimental stimulation, revealing regulatory effects not detectable in the baseline state [55, 77]. The vast majority of reQTL studies to date have been conducted on immune cells. This is not only due to the abundance of immune cells

Studies focused specifically on *trans*-eQTL mapping reduce the number of tests in other ways, such as testing only significant GWAS variants for eQTLs [58].

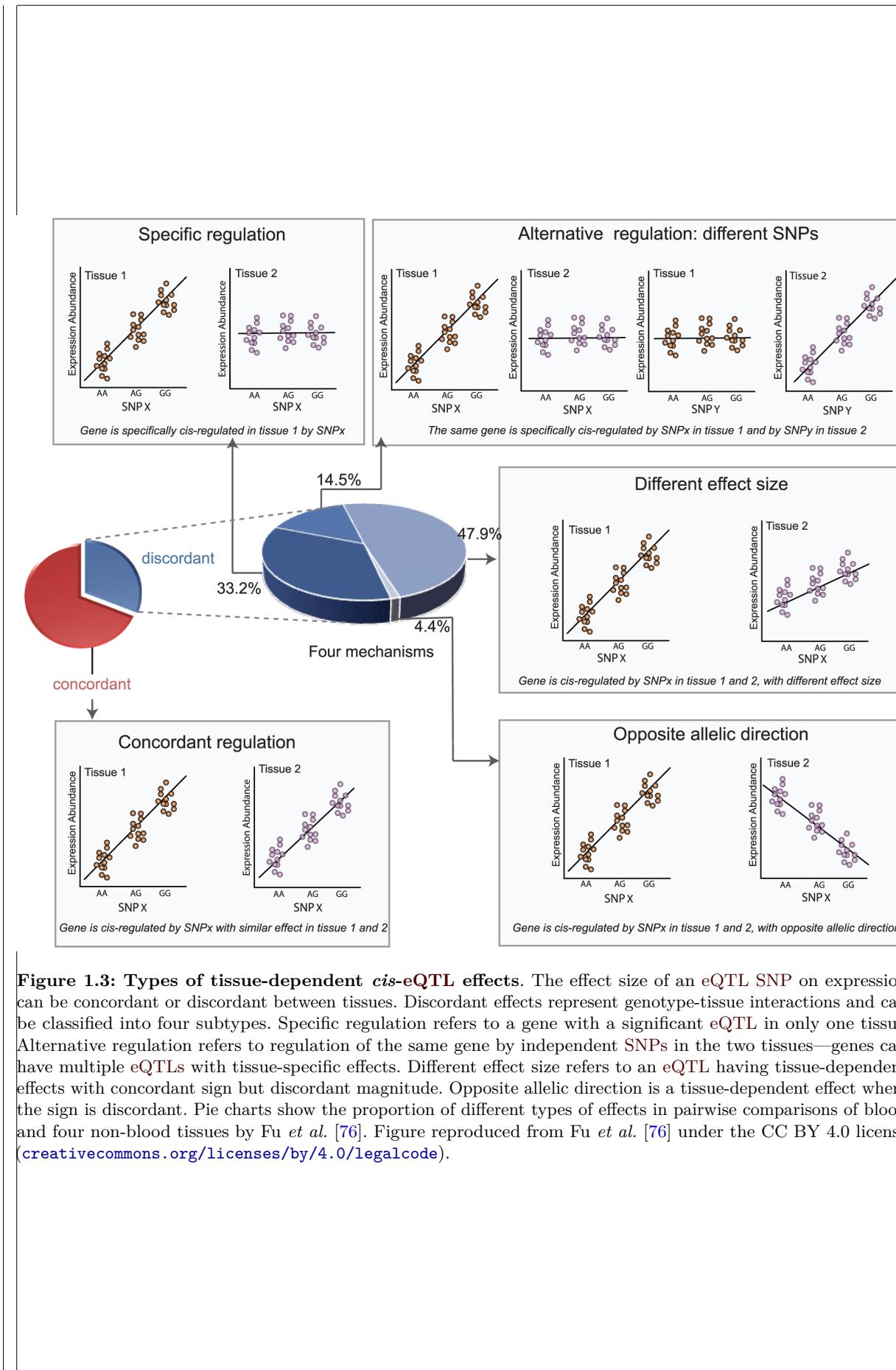


Figure 1.3: Types of tissue-dependent *cis*-eQTL effects. The effect size of an eQTL SNP on expression can be concordant or discordant between tissues. Discordant effects represent genotype-tissue interactions and can be classified into four subtypes. Specific regulation refers to a gene with a significant eQTL in only one tissue. Alternative regulation refers to regulation of the same gene by independent SNPs in the two tissues—genes can have multiple eQTLs with tissue-specific effects. Different effect size refers to an eQTL having tissue-dependent effects with concordant sign but discordant magnitude. Opposite allelic direction is a tissue-dependent effect where the sign is discordant. Pie charts show the proportion of different types of effects in pairwise comparisons of blood and four non-blood tissues by Fu *et al.* [76]. Figure reproduced from Fu *et al.* [76] under the CC BY 4.0 license (creativecommons.org/licenses/by/4.0/legalcode).

easily accessible in peripheral blood, amenable to purification and stimulation, but because the immune system is specialised for mounting different responses to different pathogens and perturbations.

When stimulation is applied *in vitro*, variables such as cell type and abundance; and the nature, length, and intensity of stimulation can be precisely controlled. A seminal study by Barreiro *et al.* [78] mapped eQTLs in monocyte-derived dendritic cells (DCs) before and after 18 h of infection with *Mycobacterium tuberculosis*. reQTLs were detected for 198 genes: 102 specific to the uninfected state and 96 specific to the infected state. They observed a 1.4-fold enrichment of reQTLs among GWAS variants associated with susceptibility to pulmonary tuberculosis, but no enrichment of eQTLs shared between uninfected and infected DCs. From overlap of reQTLs and GWAS variants, three genes (*DUSP14*, *ATP6V0A2*, *RIPK2*) were prioritised as candidates affecting tuberculosis susceptibility. Since then, numerous *in vitro* reQTL studies have been conducted with a variety of stimulations (often cytokines, pathogens, or pathogen-associated molecular patterns (PAMPs)), applied to purified [70, 79–91] or mixed cell types [83, 92].

A complementary approach is reQTL mapping with *in vivo* stimulation. An isolated mixture of cells *in vitro* cannot hope to replicate the innumerable interactions involved in human immune response. *In vivo* designs suit whole-organism stimulations and response phenotypes, such as vaccination and vaccine-induced antibody response. Published *in vivo* reQTL studies are comparatively few. Idaghdour *et al.* [93] mapped whole blood eQTLs in 94 West African children admitted to hospital for malaria and 61 age-matched controls. reQTLs with a significant case-genotype interaction were detected for five genes: *PRUNE2*, *SLC39A8*, *C3AR1*, *PADI3*, and *UNC119B*. As *SLC39A8* is upregulated with T cell activation, a postulation was made that T cell activation is important to malaria infection response. In Franco *et al.* [94], whole blood eQTLs were mapped in 247 healthy adults given trivalent inactivated influenza vaccine (TIV). Twenty genes involved in membrane trafficking and antigen processing were prioritised to be important to vaccine response, on account of having post-vaccination reQTLs or differential expression, and an expression correlation with antibody response. Lareau *et al.* [95] focused on epistatic effects of SNP-SNP interactions on expression fold-change after smallpox vaccination in 183 individuals. Eleven significant interactions were found where the effect of two independent SNPs on expression was non-additive. Apoptosis-related genes (e.g. *TRAPPC4*, *ITK*) were enriched among target genes. Most recently, Davenport *et al.* [96] mapped whole blood eQTLs in 157 systemic lupus erythematosus (SLE) patients in a phase II clinical trial of an anti-IL-6 monoclonal antibody. Nine reQTLs with effect sizes magnified by drug exposure were found to disrupt the binding site of IRF4, highlighting it as a key regulatory factor downstream of IL-6. Overall, *in vivo* reQTL studies have delivered insight into the biology of a diverse set of whole-organism phenotypes. However, ethical requirements can limit sample size and choice of stimulation. Many environmental factors (e.g. diet, lifestyle, immune exposures) cannot be controlled, potentially leading to greater experimental noise compared to *in vitro* designs, and complicating interpretation of results.

1.2.5 Gene prioritisation using eQTLs

eQTLs are enormously valuable for target gene prioritisation after GWAS. They propose both target gene and mechanism of action, where the effect of variant on complex trait is mediated through expression. GWAS variants for many traits are indeed enriched for eQTLs [97], but care must be taken when conducting enrichment analyses to avoid false positives due to the abundance of eQTLs. At current sample sizes, 60–80 % of genes have at least one detectable eQTL [55, 58]. Assuming that a locus is associated with both a trait of interest and with expression of a particular gene, how can one separate the scenario where the same causal variants affect both trait and expression (pleiotropy) from coincidental overlap between distinct sets of causal variants that may possibly be in LD? Bayesian colocalisation methods address this problem by extending Bayesian fine-mapping methods to multiple phenotypes [98–100]. Using information from all measured variants in the locus, they estimate the posterior probability that the same causal variants are associated with both phenotypes, distinguishing pleiotropy from LD.

Given the effect of an eQTL can be starkly context-dependent, eQTL datasets from trait-relevant contexts are most useful for gene prioritisation. For instance, immune *in vitro* reQTLs are enriched more so than non-reQTLs among GWAS associations for immune-related phenotypes, such as susceptibility to infectious [78, 92] and immune-mediated diseases [85, 92]. Supplementing shared eQTL effects with cell type-specific eQTL effects finds many additional colocalisations with complex traits [74, 101]. The increasing number of context-dependent eQTL datasets available for large-scale colocalisation analyses means eQTLs can propose not just target gene and mechanism, but also the specific environments most relevant to a trait.

1.3 Phenotypes of immune response

1.3.1 An overview of the immune system

Immunology began as the study of host defense against infection [102]. It is now recognised that the immune system is also involved in pathogenesis of diverse conditions encompassing allergic diseases, autoimmune and immune-mediated diseases, and cancer. This subsection provides a basic overview of parts of the immune system relevant to this thesis.

The two major arms of the immune response are the innate and adaptive response. The innate response is rapid and non-specific, occurring in the first few minutes to days after the initial (primary) exposure to infection. This triggers the adaptive response, which takes days to weeks to develop, but delivers a powerful and specific response capable of eliminating pathogens that have evaded the innate response. The adaptive response can also create immunological memory lasting years to decades, where re-exposure to the same pathogen induces a faster and more powerful recall response*. Both arms distinguish self from non-self through complex interactions between many cell types via surface receptors and signalling molecules.

Immune cell types differentiate from common myeloid progenitor or common lymphoid progenitors, which themselves are descended from pluripotent hematopoietic stem cells (HSCs) in the bone marrow. By and large, the cells of the innate response are of the myeloid lineage,

*There is increasing evidence the innate immune system also has a form of immunological memory [103].

and the cells of the adaptive response are of the lymphoid lineage. Immune cells are also called leukocytes or white-blood cells as many types can be found in peripheral blood, but certain types are confined to tissues or parts of the lymphatic system.

Innate response begins with the detection of pathogens by phagocytotic sensor cells—primarily neutrophils, tissue-resident macrophages, and DCs. These cells express pattern recognition receptors (PRRs) that recognise conserved PAMPs not present in host cells, then secrete small proteins (cytokines) that trigger the inflammatory response: a massive recruitment of multiple cell types from blood into infected tissues. Recruitment is partially mediated by a family of cytokines called chemokines, which chemically attract immune cells by creating a concentration gradient (chemotaxis). Recruited neutrophils clear pathogens by phagocytosis and secrete antimicrobial molecules by degranulation. Natural killer (NK) cells detect and kill virus-infected and tumour cells. Circulating monocytes migrate to the site of infection and differentiate into macrophages and DCs. Macrophages perform phagocytosis, modulate inflammation, and can also engage in antigen-presentation—but it is DCs that are considered to be the specialist antigen-presenting cell (APC) type. Antigen-presentation by DCs is a key link between the innate and adaptive responses.

The main forces of the adaptive response comprise B and T lymphocytes. Naive lymphocytes express antigen receptors that recognise parts of specific antigens called epitopes. When they encounter this antigen, they activate, proliferate (clonal expansion), then differentiate into effector cells. To initiate adaptive response, CD4⁺ (helper) T cells recognise peptide fragments from the antigen presented in a complex with major histocompatibility complex (MHC) class II on the surface of APCs. CD4⁺ T cells then differentiate into several subsets; these activate and regulate other immune cell types such as macrophages, CD8⁺ T cells and B cells. Activated CD8⁺ (cytotoxic) T cells recognise antigens presented by MHC class I on infected cells and directly kill the cell. Activated B cells differentiate into plasma cells that secrete large quantities of antibodies, the soluble form of the B cell receptor (BCR). Antibody-mediated immunity is also called humoral immunity, whereas T cell and innate immune responses comprise cell-mediated immunity. A small subset of activated B and T cells can become memory cells, responsible for long-term immunological memory.

1.3.2 High-throughput immunology

To understand the immune system and its intricate interactions, “systems immunology” studies take a holistic rather than reductionist experimental approach [104–106]. The basic principle is the same: experimentally perturb the immune system and observe its response. Drugs and vaccines can be used as safe and synchronised perturbations—one of the largest subfields of systems immunology is systems vaccinology, which I review in Section 2.1.4. A range of high-throughput technologies are applied to measure response at many layers of the immune system (Fig. 1.4). Longitudinal designs are common, aiming to sample timepoints corresponding to baseline, innate, and adaptive immunity. The complexity of the immune response presents a major challenge, with the richness of sampling required often restricting the sample sizes of systems immunology studies due to cost and logistics.

There are three major themes to systems immunology. Initial studies of immune response to

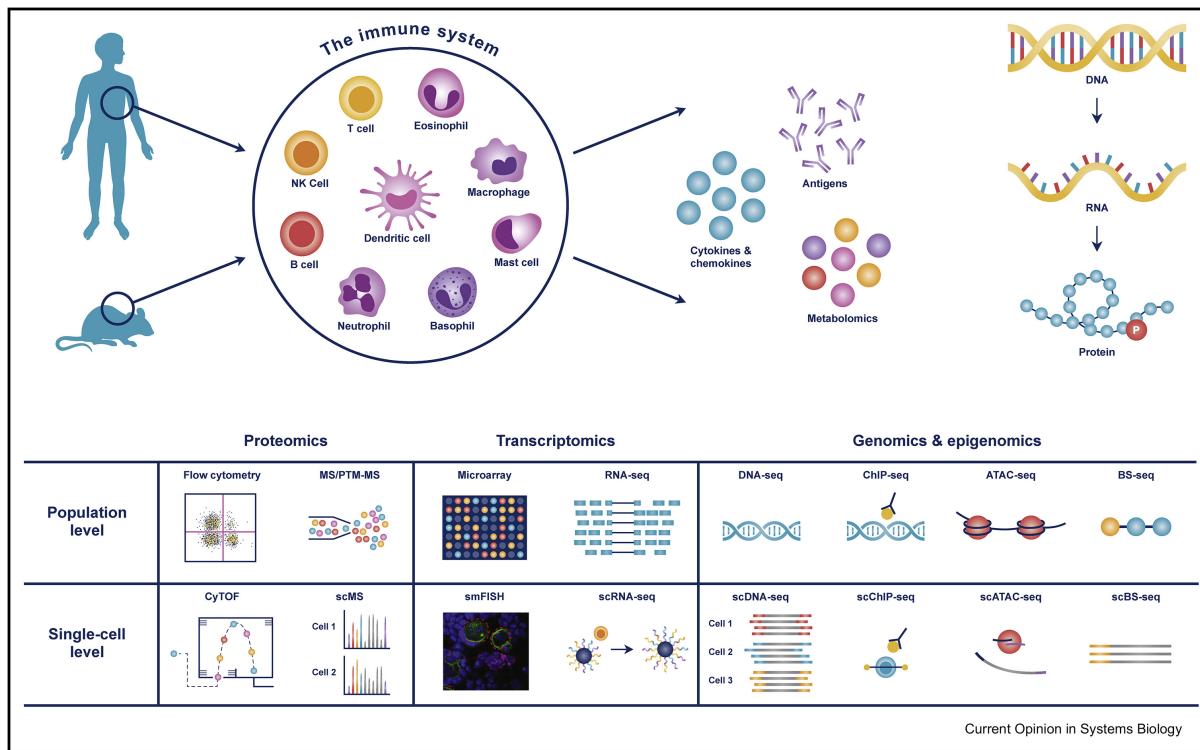


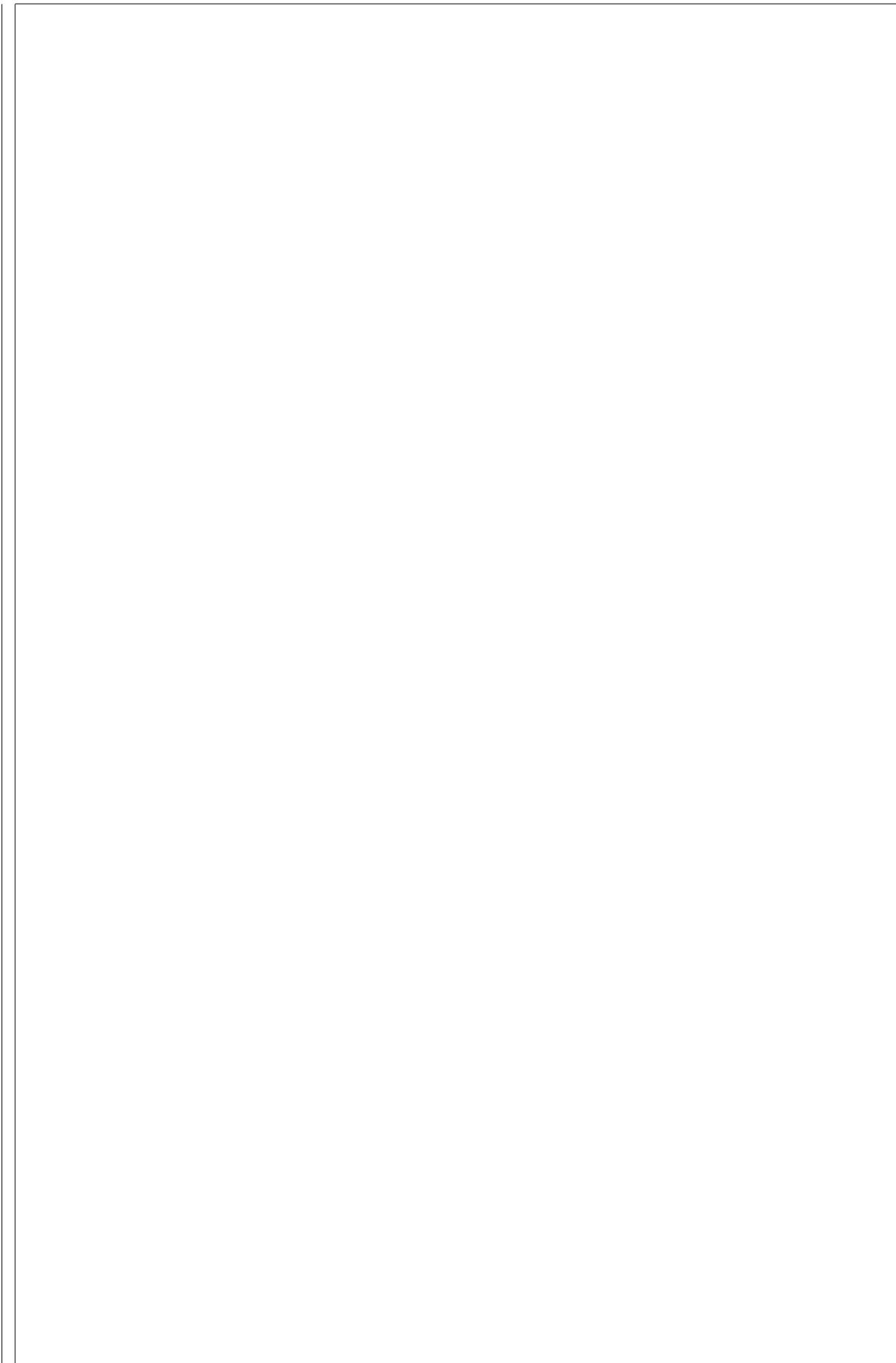
Figure 1.4: High-throughput technologies for systems immunology. Profiling can be done in humans and model organisms, at multiple levels of the immune system, and at bulk or single-cell resolutions. An additional dimension not shown here is profiling at multiple timepoints before and after perturbation. Figure reprinted from Yu *et al.* [107], © 2019, with permission from Elsevier.

a particular perturbation are often descriptive, aiming to find correlations between components of the immune response. Predictive studies then evaluate the ability to use measurements of relevant components to predict individual responses to the perturbation. Feature sets that are molecular phenotypes (e.g. gene expression) with validated predictive accuracy are known as molecular signatures. Causal inference is the third and most difficult goal. Fortunately, the heritability of immune parameters (e.g. cell counts, surface marker expression, serum protein levels) is substantial (20–40 % [108–111]), with greater heritability for innate than adaptive immune parameters [110]. Much akin to GWAS and quantitative trait locus (QTL) studies, to identify causal links in the immune system, one can leverage genetic variants as naturally occurring perturbations [105, 112]. Controlled variation can also be systematically generated by RNA interference or genome editing [113]. Obtaining causal understanding is essential for clinical translation, to determine the interventions that can be made to promote effective response to pathogens and vaccines, and impede pathways that lead to immune dysregulation in disease.

1.4 Thesis outline

This thesis examines longitudinal response to *in vivo* immune perturbations by vaccines and drugs. Chapter 2 is a descriptive differential gene expression (DGE) study of transcriptomic and antibody responses to pandemic influenza vaccine (Pandemrix) in the Human Immune Response Dynamics (HIRD) cohort of healthy adults. Chapter 3 integrates HIRD genotype data to map the regulation of expression response to Pandemrix using an *in vivo* reQTL study design. In

Chapter 4, I mirror the design of the previous two chapters, exploring clinical response to biologic anti-tumour necrosis factor (TNF) therapy for Crohn's disease (CD) patients in the Personalised Anti-TNF Therapy in Crohn's Disease (PANTS) cohort. Finally, Chapter 5 presents an overview of shared themes and limitations, and provides recommendations for future analyses and study designs for immune response phenotypes.



Chapter 2

Transcriptomic response to Pandemrix vaccine

The work presented in this chapter is a collaboration between the Wellcome Sanger Institute, King's College London, the Francis Crick Institute, and the Biomedical Research Centre at Guy's and St Thomas' Hospital and King's College London. I would like to thank Adrian Hayday, for kindly extending the opportunity to collaborate on the HIRD cohort; Efstathios Theodoridis, for performing the RNA and DNA extractions; Sean O'Farrell and Anna Lorenc, for providing the HIRD clinical, FACS, and antibody titre data, and for providing advice on the data formats; and the Wellcome Sanger Institute Sample Management and Pipelines teams, for performing the RNA-seq library preparation and sequencing, and the array genotyping.

2.1 Introduction

2.1.1 Influenza

Influenza is an infectious respiratory disease caused by the influenza virus family (*Orthomyxoviridae*) in a variety of vertebrate hosts. Of the four virus types (A, B, C, D) defined by antigenic specificity of the viral nucleoprotein, human infections are primarily caused by influenza A and influenza B. Each year, seasonal epidemics result in ~1 billion infections and 300 000–500 000 deaths worldwide. Peak seasonality is defined by low humidity, low temperature, and other climate factors. Risk factors for severe illness and death include extremes of age (infants <1 yr, elderly >65 yr), pregnancy, obesity, chronic illness, and host genetics (e.g. mutations in *IFITM3* and *IRF7*) [114, 115].

Influenza viruses are enveloped viruses with a negative-sense single-stranded RNA genome divided into segments (eight segments in influenza A and B), each encoding one or more viral proteins. Two glycoproteins occurring on the surface of the viral envelope are the main antigens targeted by the host immune system. **Haemagglutinin (HA)**, with its characteristic head-stalk structure, facilitates viral entry by binding sialic acid-containing surface receptors on host cells. **Neuraminidase (NA)** facilitates viral release, cleaving sialic acids to prevent newly-synthesised viruses aggregating to each other—viral proteins can be sialylated post-translation—and to the dying host cell in the final stages of the viral life cycle. The gradual accumulation of mutations

in these surface protein genes is known as antigenic drift, and can lead to evasion of antibody-mediated immunity acquired during previous exposures. As the virus type with the greatest prevalence, host range, and genetic diversity, influenza A is classified into a number of subtypes based on the antigenic properties of its **HA** and **NA**. At least 18 **HA** subtypes and 11 **NA** subtypes exist [116]. Although these **HA** and **NA** subtypes are all antigenically-dissimilar, there can still be cross-reactivity between subtypes, and considerable antigenic drift within subtypes [117]. Influenza B viruses are less diverse, classified into two antigenically-distinct lineages: Victoria-like and Yamagata-like [114].

On occasion, reassortment of genome segments between viruses infecting the same cell can quickly generate new strains (antigenic shift). Antigenic shifts are associated with pandemics due to lack of pre-existing population immunity [114]. Pandemics have occurred four times in modern history: 1918 (“Spanish”), 1957 (“Asian”), 1968 (“Hong Kong”), and 2009 (“swine”). Each was caused by influenza A, involving either reassortment of human and animal strains or zoonotic transmission of animal strains [118]. For instance, the 2009 pandemic was due to an influenza A strain with **HA** subtype 1 and **NA** subtype 1 gene segments of swine origin [119]: A(H1N1)pdm09*. Pandemic strains tend to enter seasonal circulation post-outbreak, potentially replacing previously-circulating strains; A(H1N1)pdm09-like strains are now the predominant seasonal A(H1N1) strain [114].

2.1.2 Seasonal influenza vaccines

Vaccination is the primary method for prevention and control of influenza. Antigenic drift and decline of vaccine-induced immunity over time means annual vaccination is recommended. Seasonal vaccines are multivalent, usually formulated against three (trivalent) or four (quadrivalent) influenza strains anticipated to circulate in upcoming influenza seasons. The **World Health Organization (WHO)**-run Global Influenza Surveillance Response System (GISRS) makes recommendations on the most representative strains for the Northern and Southern hemispheres each year, about six months before the start of the respective seasons.

There are three classes of licensed vaccines against seasonal influenza: **inactivated influenza vaccines (IIVs)**, **live attenuated influenza vaccines (LAIVs)**, and **recombinant HA vaccines** [116, 121]. **IIVs** can be split virion, containing virions disrupted with detergent, or subunit, containing further purified **HA** protein. **LAIVs** contain low-virulence, cold-adapted viruses that replicate well only in the cool upper respiratory tract. Recombinant **HA** vaccines contain purified recombinant **HA** expressed in insect cell lines rather than relying on traditional viral propagation in embryonated chicken eggs; cell-based **IIVs** are also available. Cell-based vaccines are faster to manufacture in pandemic situations, not dependent on egg supply, and avoids egg-adaptation: mismatches between vaccine and circulating strains caused by adaptation to growth in eggs.

Licensed seasonal vaccines are effective and well-tolerated in healthy adults, but particular subclasses of vaccine are recommended for different demographics [122–125]. **LAIVs** are delivered via nasal spray and are more effective than **IIVs** at mitigating transmission. They are recommended for children—the major drivers of transmission due to high viral loads and prolonged shedding [114].

*The suffix “pdm09” distinguishes the 2009 pandemic strain from the circulating seasonal A(H1N1) strains at that time [120].

[122]—but are contraindicated in young children <2 yr due to increased risk of wheezing, and also in immunocompromised individuals. Trials suggest LAIV has superior efficacy compared to IIVs in children. High-dose and adjuvanted IIV vaccines are recommended to enhance immunogenicity in the elderly. Cell-based and egg-free vaccines are suitable for people with egg allergies. No vaccines are licensed for use in infants <6 mo, but passive immunity can be conferred through vaccinating the mother.

Point estimates of seasonal vaccine efficacy range from 50–90 % in healthy adults in controlled trials. Real-world effectiveness can be as low as 10 %, depending greatly on vaccine class, choice of endpoint, the match between vaccine and circulating strains, and various host factors [115, 126]. In general, efficacy is comparable or better in children versus young adults, and lowest in the elderly due to immunosenescence. Females mount higher antibody responses than males to IIVs regardless of age, potentially mediated by sex steroid levels [115, 127]. Immune history has a major impact on vaccine response due to immune memory. Adults primed by past exposures to seasonal influenza strains have qualitatively different responses to unprimed adults or influenza-naïve children. For example, influenza-naïve children mount much higher serum antibody responses to seasonal LAIV than primed adults [123]; and antibody responses to IIV peak later in unprimed individuals, requiring two doses to generate optimal concentrations [122]. Immune history also affects response via antigenic seniority (a.k.a. immune imprinting), where the antibody response is biased towards recall against strains encountered in early childhood over generation of a *de novo* response. This is beneficial if strains with the same epitopes come back into circulation, and harmful against strains still similar enough to trigger immune memory, but with drifted epitopes [115, 128]. Finally, host genetic variation in cytokine genes, immunoglobulin genes, and the human leukocyte antigen (HLA) region are associated with antibody responses—reviewed in Section 3.1.1.

2.1.3 Quantifying immune response to influenza vaccines

The efficacy of IIVs is mostly mediated by induction of strain-specific anti-HA antibodies, although other antibodies (e.g. anti-NA) may also contribute in the case of non-purified vaccines. Antibody-secreting cells (ASCs) in peripheral blood peak around one week after vaccination, and serum antibodies peak around two to four weeks after vaccination. Antibody-mediated protection may last up to a year in healthy adults [122, 129]. The immunodominance of the HA head over the stalk means most anti-HA antibodies have epitopes in the head domain. Unsurprisingly, the resulting immune selection pressure concentrates antigenic drift in the head domain. The stalk domain is relatively conserved, hence anti-stalk antibodies are more likely to be broadly neutralising antibodies effective against multiple virus subtypes (heterosubtypic immunity) [130].

The haemagglutination inhibition (HAI) assay is an inexpensive method for quantifying serum anti-HA antibody concentrations. A serial dilution of serum is created and mixed with standardised concentrations of red blood cells (RBCs) and influenza virus. Without the presence of antibodies, the receptor site on the HA head binds to membrane-bound sialic acid on RBCs, agglutinating them into a lattice that appears as a cloudy red solution. Anti-HA antibodies inhibit agglutination, allowing the RBCs to settle, creating a clear solution with a dark red pellet. The titre value comes from the most dilute concentration of serum that completely inhibits

agglutination [131]. The value is relative to the concentrations of reagents, requiring standardised protocols for comparability. A standardised HAI titre of 40 (1:40 dilution) is deemed seroprotective, and is an accepted correlate of protection for IIVs, representing 50% clinical protection rate against infection [122, 132]. Reliable correlates of protection are useful in vaccine trials to reduce resource requirements (e.g. time, sample size, cost) compared to disease or infection-based endpoints like clinical protection [133]. For seasonal IIVs, regulatory agencies define target criteria based on the minimum proportion of individuals achieving HAI seroprotection (≥ 40 titre) and seroconversion (≥ 4 -fold increase in titre after vaccination, indicating the vaccine is immunogenic) [116, 130, 132].

An alternative method is the microneutralisation (MN) assay, which quantifies concentrations of serum antibodies capable of neutralising viral infectivity. Neutralising antibodies may be anti-HA antibodies quantifiable by HAI, but may also be anti-HA stalk antibodies or antibodies with non-HA targets not detectable by HAI [116]. The assay again involves a serial dilution of serum, which is incubated with standardised concentrations of virus. The serum-virus mixtures are inoculated into host cells *in vitro*. After incubation, virus-infected cells are quantified (e.g. enzyme-linked immunosorbent assay (ELISA) using antibodies against viral proteins), the lack of which indicates neutralising activity sufficient to suppress viral replication [131]. A MN assay value of 160 (1:160 dilution) is considered equivalent to the seroprotective HAI value of 40 [122].

IIVs primarily induce serum antibodies of the IgG isotype. The cellular response has not been extensively studied, but the induction of CD8⁺ T cells by unadjuvanted subunit IIVs is considered poor [122, 134]. In contrast, LAIVs can induce serum IgG, but also efficiently induce mucosal IgA and T cell responses [123]. Protection may also have greater duration than that afforded by IIVs, although the longevity still pales in comparison to natural infection, which can grant strain-specific protection that is lifelong [116, 121–123]. Different facets of response play different roles in immunity: serum IgG is important for limiting severity of systemic infection, mucosal IgA in the upper respiratory tract inhibits initial infection and transmission, CD8⁺ T cells promote viral clearance and recovery, and CD4⁺ T cells help induce the humoral and CD8⁺ T cell responses [114, 122, 130, 135]. Correlates of protection for LAIV have not yet been defined; licensed LAIVs have all been licensed on the basis of clinical protection. Their comparable efficacy to IIVs in adults despite low HAI titres and seroconversion rates are presumed to be mediated by mucosal and cell-mediated immunity [116, 123]. Clearly, a broader view of immunity than granted by serological antibody assays is needed to understand the mechanisms leading to efficacious influenza vaccine responses.

2.1.4 Systems vaccinology of seasonal influenza vaccines

Vaccinology has historically been driven by the “isolate-inactivate-inject” paradigm [136]. Many vaccines have been developed and licensed through expensive, large-scale, and largely empirical trials that deliver highly effective vaccines, but little understanding of the immunological mechanisms of protection. In response, the last decade has seen the rise of systems vaccinology, a subfield of systems immunology dedicated to the analysis of high-throughput data measured at multiple levels of the immune system to characterise response to vaccination [133, 137–143]. Traditional serological assays (e.g. HAI, MN) are complemented with a raft of other technologies

to give a broader view of immune response [137–139, 142, 143]. Flow (e.g. fluorescence-activated cell sorting (FACS)) and mass cytometry (e.g. cytometry by time-of-flight (CyTOF)) are used to quantify immune cell subpopulations by their surface markers using fluorescent and heavy metal tags. These technologies can also be used to quantify intracellular markers (e.g. cytokines) by cell staining. Frequencies of cells secreting specific proteins (e.g. cytokines or antibodies) can also be quantified (e.g. enzyme-linked immune absorbent spot (ELISPOT)), useful for monitoring activated cell populations involved in both humoral and cell-mediated immunity. The transcriptome of peripheral blood is extremely popular to assay (e.g. expression array, RNA sequencing (RNA-seq)), providing an accessible, global measure of gene expression in dozens of immune cell subtypes without the need to select specific genes of interest in advance. Recently, there has been a growing interest in targeted sequencing of B cell and T cell repertoires, responsible for the specificity of the adaptive immune system. Serum proteins can be quantified in a low-throughput (e.g. ELISA) or multiplex manner (e.g. Luminex). Modern proteomics platforms also embrace a global philosophy, simultaneously quantifying thousands of proteins (e.g. SOMAscan). Finally, although not often considered due to small cohort sizes, host genetic variation can be accurately measured by genotyping arrays and sequencing.

Longitudinal study design is key, not only to profile different stages of innate and adaptive immunity, but also for determining correlates of protection. Correlates are known for some but not all established vaccines [144, 145]. For novel and emerging diseases, there may be no prior knowledge of correlates for use in vaccine development. The term “molecular signature” was coined to refer to transcriptomic responses induced early after vaccination that correlate with, and importantly, are predictive of later immune phenotypes (e.g. antibody titres) [133], although non-transcriptomic data types can also be used to form signatures. The ultimate goal is baseline prediction, where the immune state immediately prior to vaccination predicts response, and could potentially be modulated to enhance response in a similar manner to adjuvanting the vaccine itself [146].

Work in the field has thus far focused on established vaccines. One can learn from the success of highly-efficacious vaccines like yellow fever vaccine (YF-17D); where interferon, complement, and inflammasome expression signatures measured 3–7 days post-vaccination predict CD8⁺ T cell and neutralising antibody responses 60 days post-vaccination [138, 147]. Much has also been learnt from the study of vaccines with suboptimal efficacy in challenging populations: infants and the elderly, pregnant women, immunocompromised patients, ethnically-diverse populations, developing countries [148]. The field has not yet identified completely novel correlates for many vaccines, partially because protection itself can be difficult to measure. One promising system is the human challenge trial, applied by Vahey *et al.* [149] to identify genes in the immunoproteasome pathway associated with protection from malaria challenge after adjuvanted RTS,S malaria vaccination. If correlates for novel vaccine candidates could be routinely established based on shared immune mechanisms leading to efficacy and long-lasting protection derived from multiple successful vaccines, there is enormous potential for optimising trials to be fast and cost-effective [133, 150], and informing rational, mechanism-based design for diseases that have thus far proved intractable to empirically-designed vaccines (e.g. HIV, malaria, non-childhood tuberculosis) [136, 140, 143, 150].

Seasonal influenza vaccines have been well-studied by systems approaches. One of the earliest studies by Zhu *et al.* [151] found that expression of type I interferon-modulated genes at day 7 was more prominent for LAIV than trivalent inactivated influenza vaccine (TIV) in children (total cohort size $n = 85$). A subsequent study found that both LAIV and TIV could induce interferon-related genes in children ($n = 37$), but much earlier in TIV (day 1) than in LAIV (day 7) [152]. As serum antibody titres are an established correlate for TIV, studies have been carried out to identify its molecular basis. Bucasas *et al.* [153] ($n = 119$) reported a 494-gene expression signature (including *STAT1*, *CD74*, and *E2F2*) measured at day 1 and 3 that correlated with serum antibody titres measured 14 and 28 days after vaccination. Signatures including day 3 kinase *CaMKIV* expression were found to predict day 28 HAI antibody titres in independent trials over three consecutive influenza seasons ($n = 67$) [154]. Expression of gene sets related to B cell proliferation at day 7 were likewise predictive of day 28 HAI ($n = 15$) [155]. Work has also been conducted to understand the heterogeneity in response due to host factors like sex [127] and age [156–159].

Signatures can be derived from predictors measured pre-vaccination [146]. A gene module enriched in apoptosis-related genes measured at baseline was found to predict day 28 HAI response ($n = 89$) [160]. Tsang *et al.* [161] used not gene expression, but FACS measurements to establish signatures for day 70 neutralising antibody titres ($n = 63$). Frequencies of several B cell, myeloid dendritic cell (DC), CD4⁺ memory T cell, and a number of other activated T cell populations were not only predictive, but also stable over a period of two months. Nakaya *et al.* [157] used data collected over five consecutive seasons ($n = 212$) to identify associations between day 28 HAI and baseline expression modules annotated to B cells (positive association), T cells (positive association), and monocytes (negative association). They were able to replicate these associations using published data from Franco *et al.* [94] and Furman *et al.* [160]. Another multi-cohort, multi-season study ($n > 500$) by the Human Immunology Project Consortium (HIPC) [159] found baseline expression of genes (*RAB24*, *GRB2*, *DPP3*, *ACTB*, *MVP*, *DPP7*, *ARPC4*, *PLEKHB2*, *ARRB1*) and gene modules to be associated with antibody response in young individuals. Again, the authors were able to validate the associations in an independent cohort.

To conclude, it must be noted that the utility of molecular signature for predicting response to influenza or other vaccines in clinical trials has not yet been validated, and it is difficult to draw causal insights from studies that are largely descriptive or predictive. The existence of temporally-stable and replicable signatures is, however, encouraging.

2.1.5 The Human Immune Response Dynamics (HIRD) cohort

For studies of seasonal influenza vaccines in adults, responses are heavily influenced by immunological memory built by past vaccination or infection with antigenically-similar strains [137, 162]. Dependence on exposure is reflected in high variability of baseline vaccine-specific antibody titres and memory B cell numbers [161]. There have also been few systems vaccinology studies of adjuvanted influenza vaccines, known to have greater immunogenicity and efficacy than non-adjuvanted vaccines in children and the elderly [158, 163, 164]. The Human Immune Response Dynamics (HIRD) study conducted by Sobolev *et al.* [162] was conceived as a unique

opportunity to study response to an adjuvanted pandemic influenza vaccine (Pandemrix), where responses are more likely to be primary than recall, and variability due to prior exposure is minimised.

Pandemrix was one of several vaccines rapidly developed and licensed in response to the 2009 influenza pandemic [165]. It is a monovalent split-virion **IV** against the pandemic influenza A/California/07/2009 (H1N1)pdm09 strain* developed by GlaxoSmithKline, containing 3.75 µg **HA** and adjuvant AS03 (oil-in-water emulsion containing DL- α -tocopherol, squalene, polysorbate 80). Subsequent studies estimated its effectiveness to be ~80% after a single dose [166]. As the H1N1 subtype had not circulated since the 1918 pandemic, the majority of the population was expected to be immunologically-naive at the time of study sampling (March 2010 to August 2011).

The study was a longitudinal, prospective cohort study. A total of 178 healthy adults in the UK were vaccinated with a single dose of Pandemrix. Clinical, transcriptomic, immune cell frequency, cytokine level, antibody titre, and adverse event phenotypes were collected. Genes associated with both myeloid and lymphoid effector functions had increased expression at day 1 versus baseline, most prominently for genes associated with the interferon response. Day 1 gene expression was impacted by age; significant global differences were observed in individuals older than 30–40 yr, considerably earlier than usually considered in studies of immunosenescence in old age. The early myeloid responses—increase in blood monocyte levels and cytokines associated with innate activation e.g. CCL4—were overall consistent with studies of unadjuvanted seasonal influenza vaccines. However, the early lymphoid responses—driven by a five-fold increase in serum interferon gamma levels at day 1—were unique to this adjuvanted pandemic influenza vaccine.

Vaccine (antibody) response was defined as a ≥ 4 -fold increase in either **HAI** or **MN** titres after vaccination. Genes related to plasma cell development and antibody production were more highly expressed in 23 responders compared to 18 non-responders at day 7 post-vaccination. However, due to high variability among the vaccine non-responders in expression trajectory over time, a predictive model that segregated the two groups could not be built, even considering other predictors such as frequencies of immune cell subsets, and serum cytokine levels. There appeared to be many “routes to failure” [162], rather than any single determining factor leading to poor antibody response.

2.1.6 Chapter summary

Transcriptomic measurements in the original **HIRD** study were restricted to a relatively small number of individuals ($n = 46$), limiting power to detect expression associated with antibody response. In addition, the binary responder versus non-responder definition used does not account for variation in baseline titres, and dichotomisation of a continuous variable loses information and implies a discontinuity in response at the cutoff.

In this chapter, I combine the existing array data with newly generated **RNA-seq** data ($n = 75$)

*The **WHO** nomenclature for isolates specifies influenza type (A, B, C, D), host of origin (human if omitted), geographical origin, strain number, year of isolation, and isolate subtype (combination of **HA** and **NA** subtypes) [117].

on additional individuals from the **HIRD** cohort using Bayesian random-effects meta-analysis to account for between-platform heterogeneity. I also compute a baseline-adjusted, continuous phenotype of antibody response to vaccination, the **titre response index (TRI)** [153]. Leveraging the greater sample size, more statistically efficient definition of vaccine response, and greater sensitivity of rank-based gene set enrichment analysis over per-gene analysis, I identify gene expression modules associated with magnitude of antibody response. The strongest associations are seen at day 7, but significant module associations are also observed at baseline.

2.2 Methods

2.2.1 Existing HIRD data and additional data generation

The design of the **HIRD** study is described fully in Sobolev *et al.* [162]. In brief, blood samples were collected from each individual on each of six visits: two pre-vaccination (days -7 and 0), and four post-vaccination (days 1, 7, 14 and 63). A single Pandemrix dose was administered after blood sampling on day 0. Serum antibodies were measured for all individuals ($n = 178$) on days -7 and 63 using both **HAI** and **MN** assays. **peripheral blood mononuclear cell (PBMC)** gene expression was profiled for 46 individuals by expression array on days -7, 0, 1 and 7.

In addition to this existing data, **PBMC RNA-seq** data was generated for 75 individuals at days 0, 1, and 7; and 169 individuals were genotyped. The sets of individuals with gene expression assayed by array and **RNA-seq** are disjoint, as no biological material for RNA extraction remained for the array individuals. An overview of datasets is shown in Fig. 2.1.

2.2.2 Computing baseline-adjusted measures of antibody response

There were 166/178 individuals with **HAI** and **MN** titres available at both baseline (day -7) and post-vaccination (day 63). Sobolev *et al.* [162] defined Pandemrix vaccine responders as individuals with ≥ 4 -fold titre increases from day -7 to day 63 in either the **HAI** or **MN** assays. This is a typical threshold for **HAI** and **MN** seroconversion used to assess the immunogenicity of seasonal **IIVs** [116], and has also been recommended for pandemic **IIVs** [167]. However, they noted there was “a complete spectrum” of baseline titres of non-responders, citing “glass ceiling” non-responders whose high baseline titres made “enhancements by ≥ 4 -fold harder to achieve”. This may be referring to the dynamic range of the assays. In the full data, the range of **HAI** titres is 8–4096, and the range of **MN** titres is 10–5120 (Fig. 2.2a, Fig. 2.2b). In just the day -7 baseline titres, the range of **HAI** titres is 8–512, and the range of **MN** titres is 10–5120*. It is impossible for an individual with higher than 1280 **MN** at day -7 to achieve a 4-fold increase in **MN** after vaccination if the maximum **MN** value is 5120. This ceiling effect can be seen in Fig. 2.2d, where for a given baseline **MN** titre, there is a limit to the maximum observable fold change.

Another perspective is to consider that day 63/day -7 fold change is a change score on the log scale. It is well-known that change scores are usually negatively correlated to baseline. This can

*This indicates some individuals likely do have pre-existing antibodies to the pandemic strain (or cross-reactive antibodies), although the mean of the baseline titre distribution would still be expected to be higher if this were a seasonal vaccine.

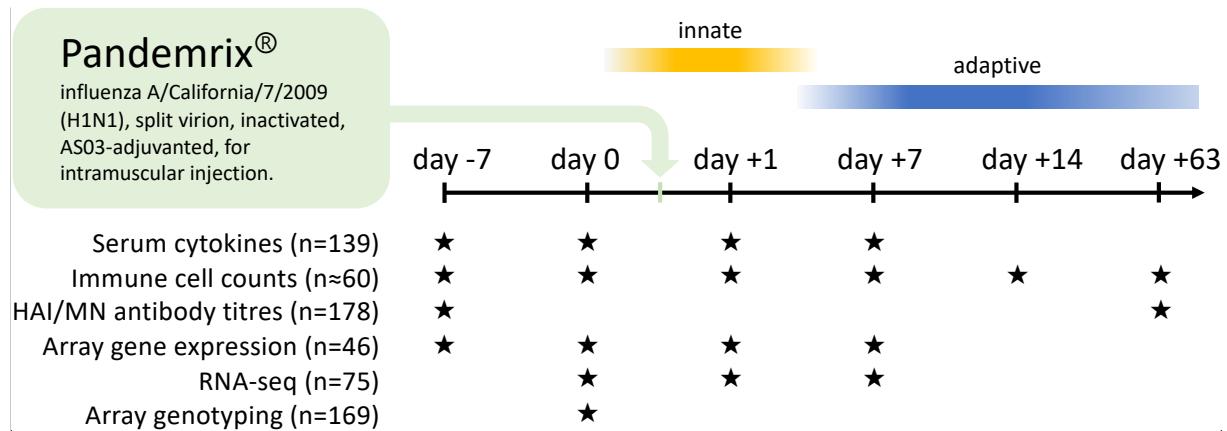


Figure 2.1: Overview of HIRD study data. The total cohort size was 178 individuals. Serum cytokines were quantified by 16-plex Luminex panel. Immune cell subsets were quantified by FACS. Serum antibodies were quantified by both HAI and MN assays. Array and RNA-seq gene expression were quantified in the PBMC compartment.

be due to individual-level regression to the mean*, the tendency for extreme observations to be followed by less extreme ones in the same individual [168], but is also due to the mathematical relationship between change score and baseline (“mathematical coupling” [170]). The correlation between change score and baseline is likely to be negative when the variance of the post-test score is much larger than the variance of the baseline and the correlation between baseline and post-test score is less than one [170, 171]. The negative correlation of titre fold change and baseline is visible in the HIRD data (Fig. 2.2c, Fig. 2.2d).

Additionally, dichotomisation of continuous variables can result in loss of information [172–175]. Cohen [172] presents a classic example where dichotomising a continuous independent variable reduces statistical power akin to throwing away a third of the samples—this being the optimal case when the cutpoint is the mean. A discontinuous cutpoint is also biologically implausible, implying that a 4.01-fold antibody titre change would be dramatically more protective than a 3.99-fold change.

To address these concerns, I computed the **TRI** as defined in Bucasas *et al.* [153]. For each assay, a linear regression was fit with the \log_2 day 63/day -7 titre fold change as the response, and the \log_2 day -7 baseline titre as the predictor. The residuals from the two regressions were each standardised to zero mean and unit variance, then averaged with equal weight. The **TRI** is a single variable that expresses a continuous measure of change in antibody titres averaged across both assays post-vaccination, compared to individuals with a similar baseline titre. It is no longer correlated with baseline (Fig. 2.2e, Fig. 2.2f), and remains qualitatively comparable to the original binary definition (Fig. 2.2g, Fig. 2.2h).

Descriptive statistics for the 114 individuals with both gene expression and antibody titre data are presented in Table 2.1. Although the proportion of responders between array (32/44) and RNA-seq (59/70) individuals is similar ($p = 0.16$, Fisher’s exact test), the variance of **TRI** in array individuals is higher ($p = 2.10 \times 10^{-4}$, Levene’s test), suggesting more extreme antibody response phenotypes are present (Fig. 2.3). The cause of this is unknown—there is a possibility

* Cf. group-level regression to the mean, which is prominent if the baseline measurement is used as a selection criteria for follow-up [168, 169].

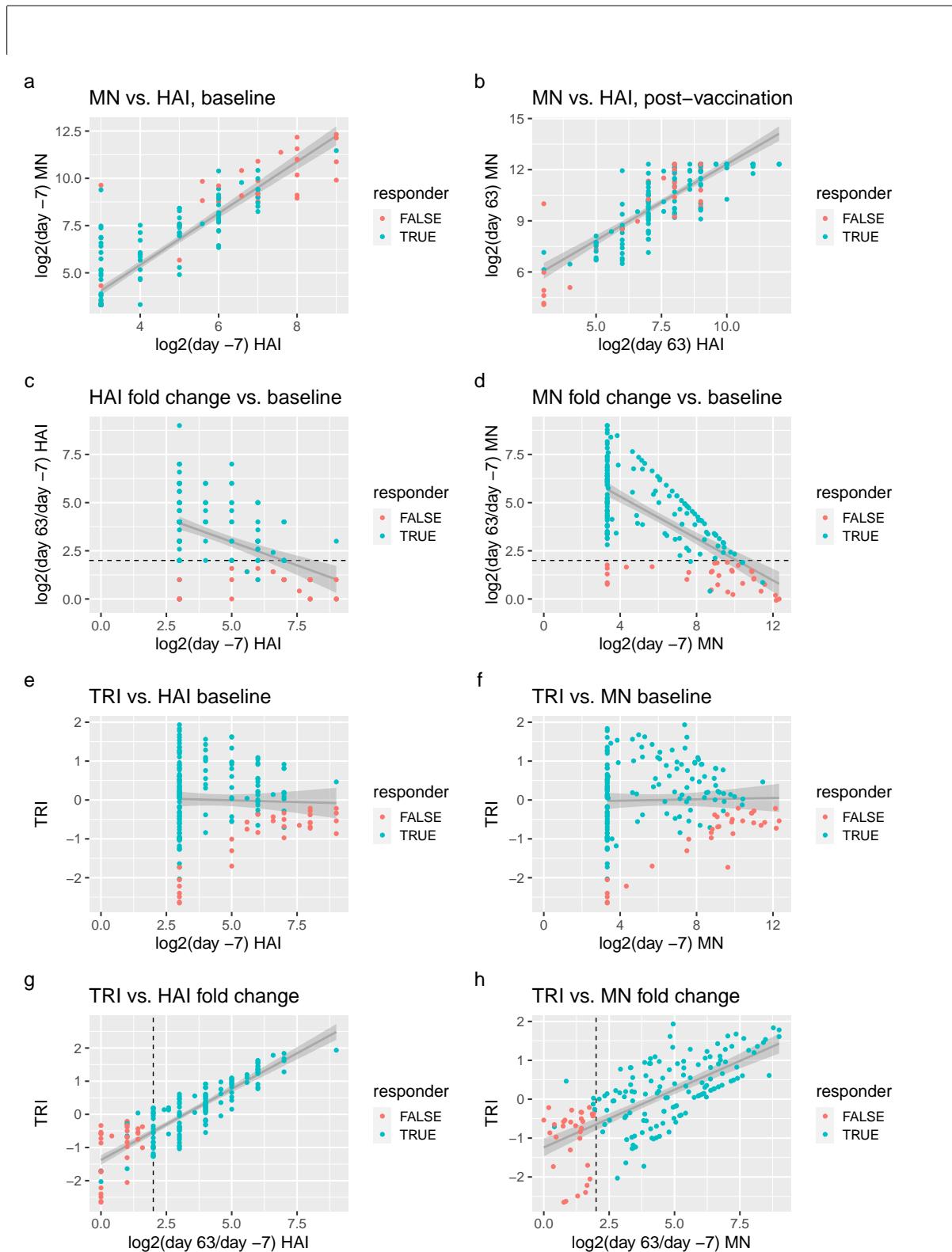


Figure 2.2: Antibody titre data and responder definitions. Titre values are on the \log_2 scale. Individuals are colored by binary responder status: ≥ 4 -fold increase in either HAI or MN titres from baseline (day -7) to post-vaccination (day 63). Dashed lines show the ≥ 4 -fold thresholds. (a, b) HAI and MN titres are correlated at baseline (a) and post-vaccination (b). (c, d) Baseline titres are negatively correlated to fold change. (e, f) TRI is computed from the standardised residuals from c and d, adjusting for baseline titre. (g, h) TRI remains comparable in ordering to binary response status.

that individuals with more extreme phenotypes were prioritised for array transcriptomics in the original HIRD study*.

2.2.3 Genotype data generation

DNA was extracted from frozen blood using the Blood and Tissue DNeasy kit (Qiagen), and genotyping was performed using the Infinium CoreExome-24 BeadChip array (Illumina). In total, 192 samples from 176 individuals in the HIRD cohort—replicate samples were submitted for individuals where extracted DNA concentrations were initially low—were genotyped at 550 601 markers

2.2.4 Genotype data preprocessing

Using PLINK (v1.90b3w) [176], genotype data underwent the following quality control filters to remove poorly genotyped samples and markers:

- maximum marker missingness across samples <5 %;
- maximum sample missingness across markers <1 % (Fig. 2.4);
- sample heterozygosity rate within 3 standard deviations of the mean of all samples (threshold selected visually to exclude outliers, Fig. 2.4);
- sample sex mismatches based on X chromosome marker heterozygosity (`--check-sex` option);
- and marker deviation from Hardy-Weinberg equilibrium (HWE), an indication of genotyping or genotype calling errors [177–179] (`--hwe` option, p -value $<1 \times 10^{-5}$)†.

To exclude closely-related individuals and deduplicate samples from the same individual, pairwise kinship coefficients were computed using KING (v1.4) [181]. As rare variation is not generally required to determine relatedness, markers were filtered to minor allele frequency (MAF) >0.05 for computational efficiency. For each pair of samples with pairwise kinship coefficient >0.18 (first-degree relatives or closer), the sample with lower marker missingness was selected. After all filters, 169/176 samples and 549 414/550 601 markers remained.

2.2.5 Computing genotype principal components as covariates for ancestry

As shown in Table 2.1, the HIRD cohort is multi-ethnic. Large-scale population structure explains variation in gene expression [182, 183], so including genotype principal components (PCs) that reflect that structure as covariates can increase statistical efficiency for detecting associations with expression. I used HapMap 3 samples [184] as a reference population of unrelated individuals where the major axes of variation in genotypes are ancestry. Genotypes were first linkage

*Personal communication with Sobolev *et al.* [162] authors.

†A wide range of thresholds for the HWE marker filter in controls between 1.00×10^{-3} and 5.70×10^{-7} are reported in the literature [178]. The HWE threshold used here is from de Lange *et al.* [180]; since the HIRD cohort is two orders of magnitude smaller in size, this represents a relaxed threshold, so additional vigilance for genotyping errors downstream is required. In principle, it may be possible to select an appropriate threshold from the empirical distribution of HWE p -values [177].

Table 2.1: Descriptive statistics for HIRD individuals with both expression and antibody data.
Values are count and percentage for categorical variables; mean and standard deviation for continuous variables.

	Total n = 114	Platform Array n = 44	Platform RNA-seq n = 70
Gender			
F	72 (63.2%)	27 (61.4%)	45 (64.3%)
M	42 (36.8%)	17 (38.6%)	25 (35.7%)
Age at vaccination (years)	29.2 (11.8)	32.9 (14.1)	26.8 (9.4)
Ancestry (self-reported)			
Asian	14 (12.3%)	5 (11.4%)	9 (12.9%)
Black/African	9 (7.9%)	4 (9.1%)	5 (7.1%)
Caucasian	82 (71.9%)	33 (75%)	49 (70%)
Latin American	2 (1.8%)	1 (2.3%)	1 (1.4%)
Mixed	5 (4.4%)	1 (2.3%)	4 (5.7%)
Other - Arab	1 (0.9%)	0 (0%)	1 (1.4%)
White Other	1 (0.9%)	0 (0%)	1 (1.4%)
log2 day -7 HAI	4.4 (1.8)	4.2 (1.6)	4.5 (1.9)
log2 day 63 HAI	7.6 (1.8)	7.4 (2.2)	7.6 (1.5)
log2 HAI fold change	3.2 (1.9)	3.2 (2.4)	3.1 (1.6)
log2 day -7 MN	6.2 (2.8)	5.4 (2.4)	6.6 (3.0)
log2 day 63 MN	10.4 (2.0)	9.5 (2.2)	10.9 (1.6)
log2 MN fold change	4.2 (2.3)	4.1 (2.6)	4.3 (2.1)
Responder (binary definition)			
FALSE	23 (20.2%)	12 (27.3%)	11 (15.7%)
TRUE	91 (79.8%)	32 (72.7%)	59 (84.3%)
TRI	-0.0 (0.9)	-0.2 (1.2)	0.1 (0.7)

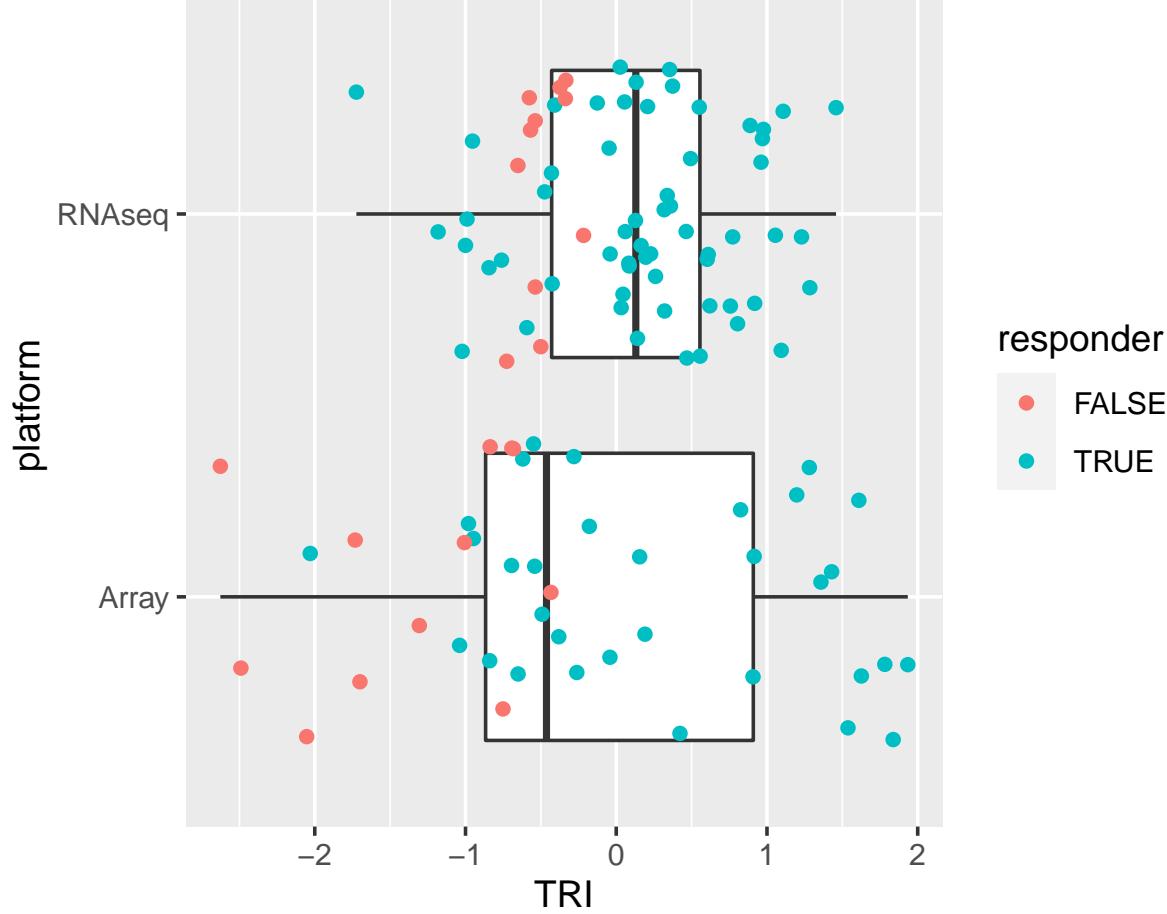


Figure 2.3: Distribution of patient TRIs, stratified by expression measurement platform. Points are colored by binary response status.

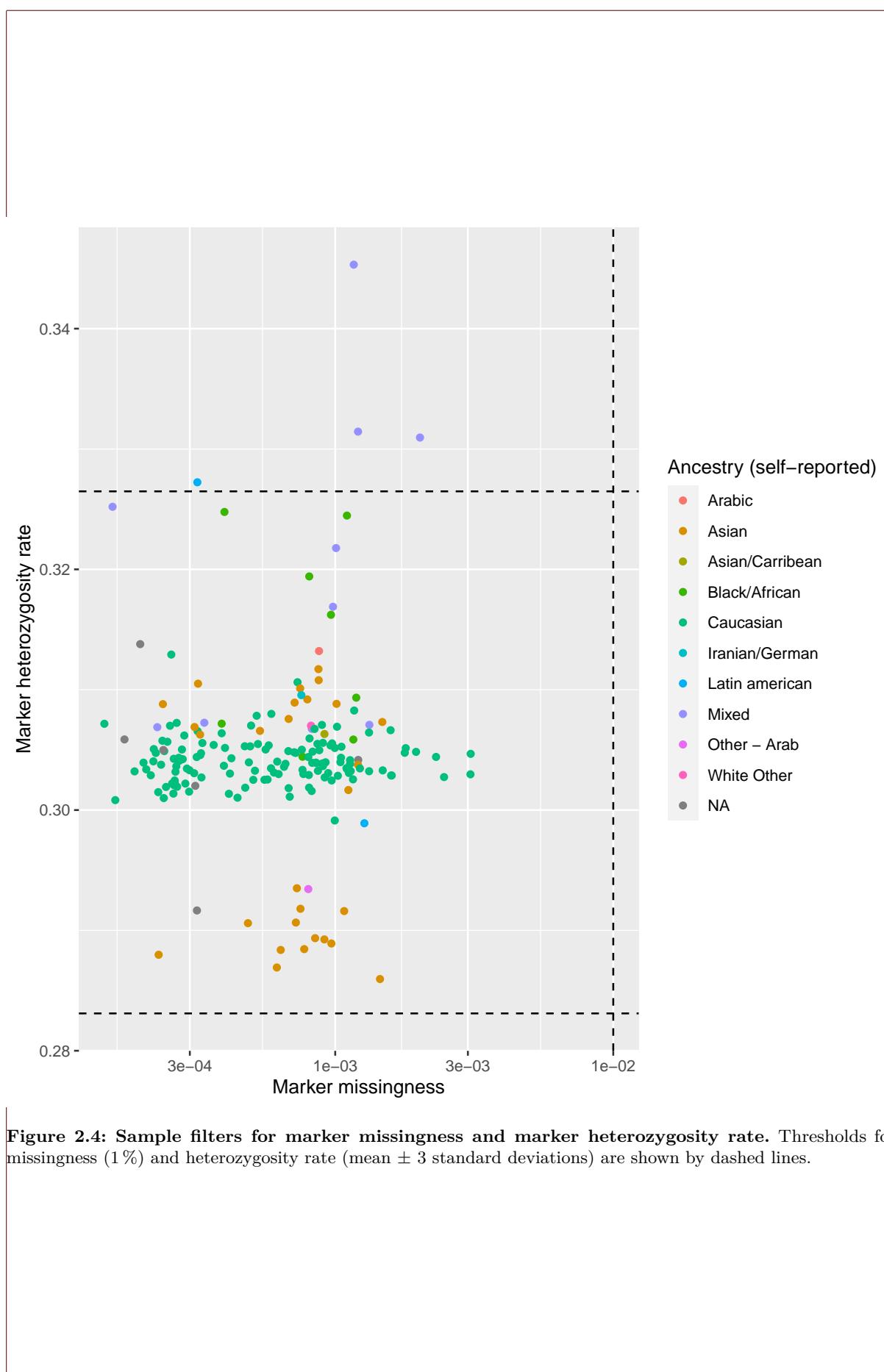


Figure 2.4: Sample filters for marker missingness and marker heterozygosity rate. Thresholds for missingness (1%) and heterozygosity rate ($\text{mean} \pm 3$ standard deviations) are shown by dashed lines.

disequilibrium (LD)-pruned (PLINK --indep-pairwise 50 5 0.2 i.e. in a sliding window of 50 kbp, with a step size of 5 variants, remove variants at each step until no pair of variants has LD >0.2), to avoid regions with many redundant markers being overrepresented in the resulting PCs [185, 186]. Eighteen genomic regions with especially strong and/or long-range LD that contain many highly correlated markers were excluded, otherwise some PCs may reflect those just regions rather than genome-wide ancestry [185, 187]. Principal component analysis (PCA) was performed using smartpca (v8000) [185], then HIRD sample PCs were computed by projection onto the HapMap 3 PCA eigenvectors. A projection was used instead of in-sample PCA, as cryptic relatedness in HIRD may be reflected in the resulting PCs instead of ancestry [188]. For non-genotyped individuals with expression data, PC values were imputed as the mean value for all genotyped individuals with the same self-reported ancestry. The top PCs indeed separate HIRD samples by ancestry (Fig. 2.5). Significant PCs with large eigenvalues unlikely to be due to sampling noise were selected by Tracy-Widom test [189]. The fourth PC had an eigenvalue of 1.01 ($p = 0.02$), so the top four PCs were retained as continuous covariates for ancestry downstream.

2.2.6 RNA-seq data generation

Total RNA was extracted from PBMCs using the Qiagen RNeasy Mini kit, with on-column DNase treatment. RNA integrity was checked on the Agilent Bioanalyzer and mRNA libraries were prepared with the KAPA Stranded mRNA-Seq Kit (KK8421), which uses poly(A) selection. To avoid confounding of timepoint and technical effects from library preparation and sequencing, samples were pooled by library preparation plate (three pools) ensuring libraries from all timepoints of an individual were in the same pool, then sequenced across multiple lanes as technical replicates (HiSeq 4000, 75 bp paired-end).

RNA-seq quality metrics were assessed using FastQC* and Qualimap [190], then visualised with MultiQC [191]. Sequence quality was high, as measured by mean per-base Phred scores across sample reads (Fig. 2.6). The unimodal GC-content distribution suggested negligible levels of non-human contamination (Fig. 2.7).

2.2.7 RNA-seq quantification and preprocessing

Reads were quantified against the Ensembl reference transcriptome (GRCh38.p15) using Salmon [192] in quasi-mapping-based mode, which internally corrects for transcript length and GC composition by computing an effective length for each transcript. Relative transcript abundances were summarised to Ensembl (release 90) gene-level count estimates using tximport (scaledTPM method, which scales Salmon transcripts per million (TPM) values up to the library size [193, 194]) to improve statistical robustness and interpretability [193]. To combine technical replicates, as the sum of Poisson distributions remains Poisson-distributed, counts for technical replicates were summed for each sample. The mean number of mapped read pairs per sample after summing was 27.09 million read pairs (range 20.24–39.14 million), representing a mean mapping rate of 80.73 % (range 75.57–90.10 %). These meet sequencing depth recommendations for differential gene expression (DGE) experiments (e.g. diminishing returns after 10 million single-end reads

*<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

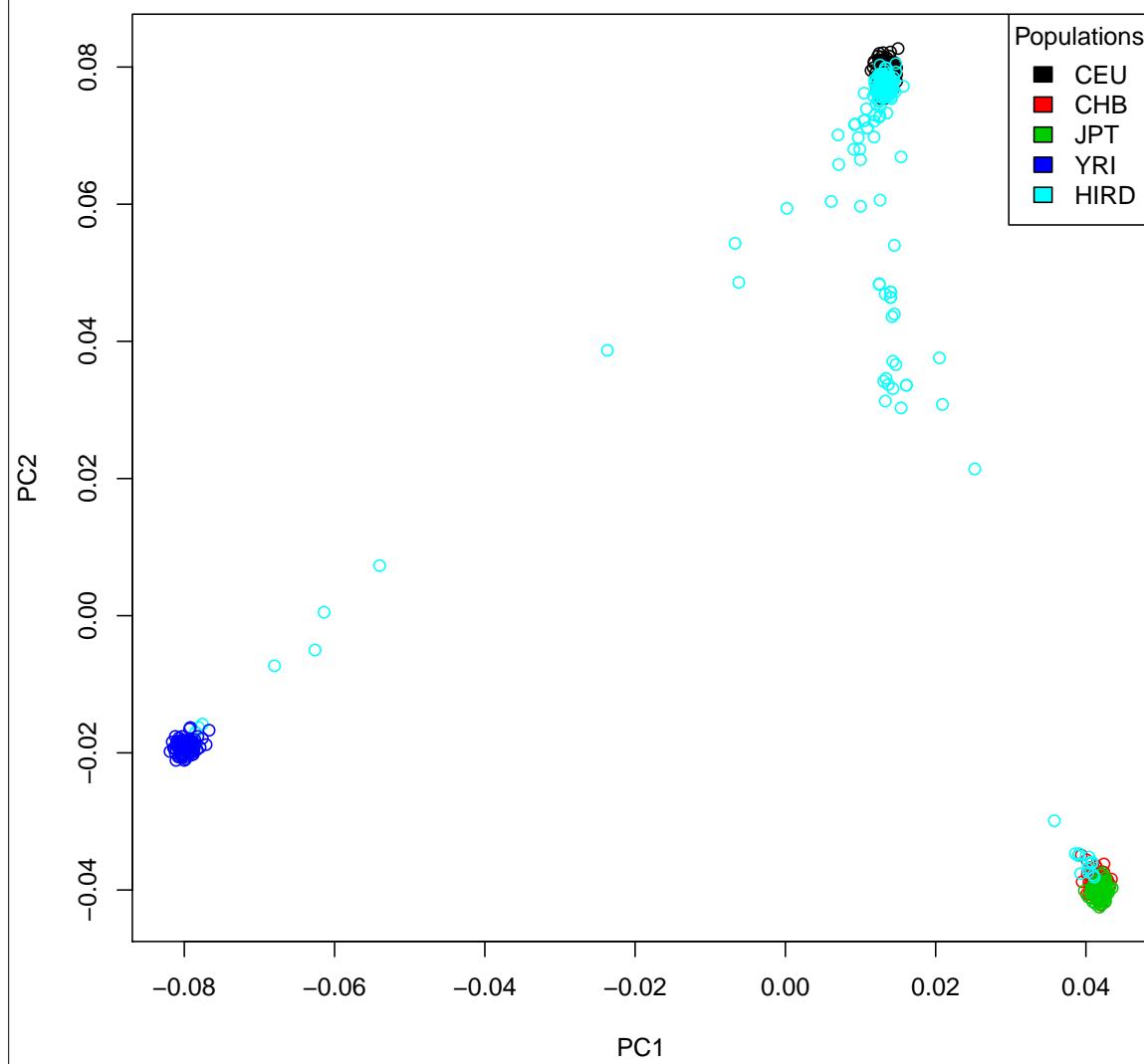


Figure 2.5: HIRD samples projected onto PC axes defined by PCA of HapMap 3 samples. The first two PCs separate individuals of European (CEU, upper-right) from Asian (CHB and JPT, lower-right) and African (YRI, lower-left).

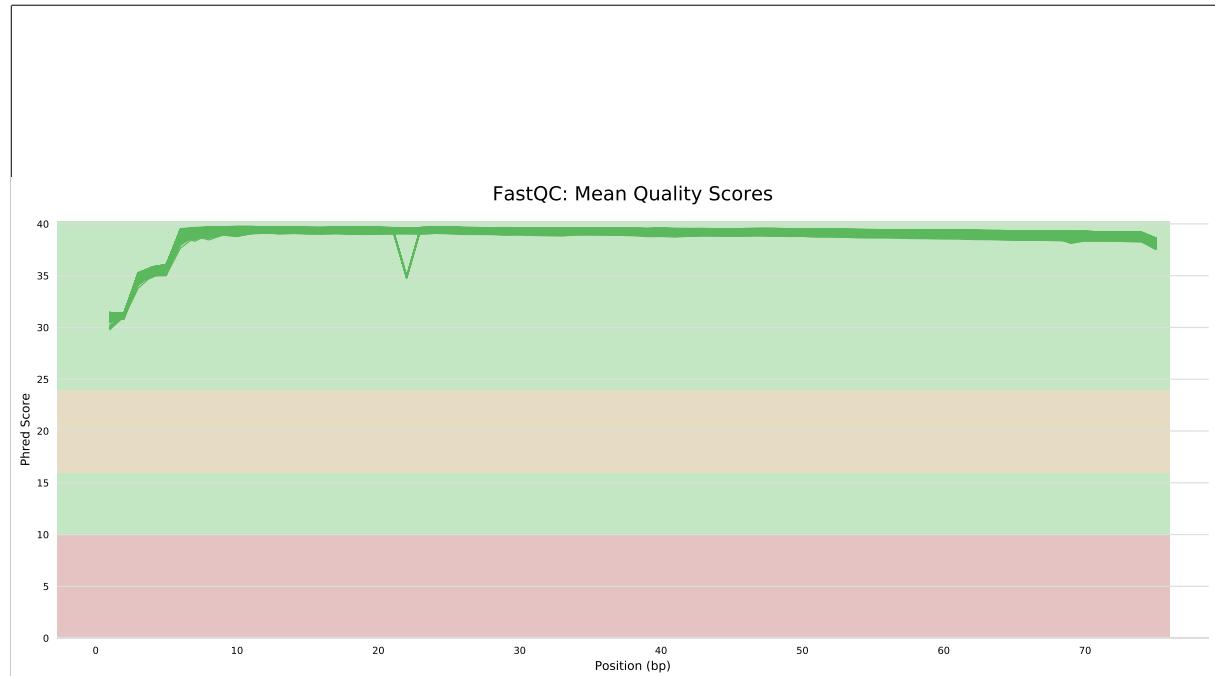


Figure 2.6: FastQC per-base sequence quality (Phred scores) versus read position for RNA-seq samples. Visualised with MultiQC [191].

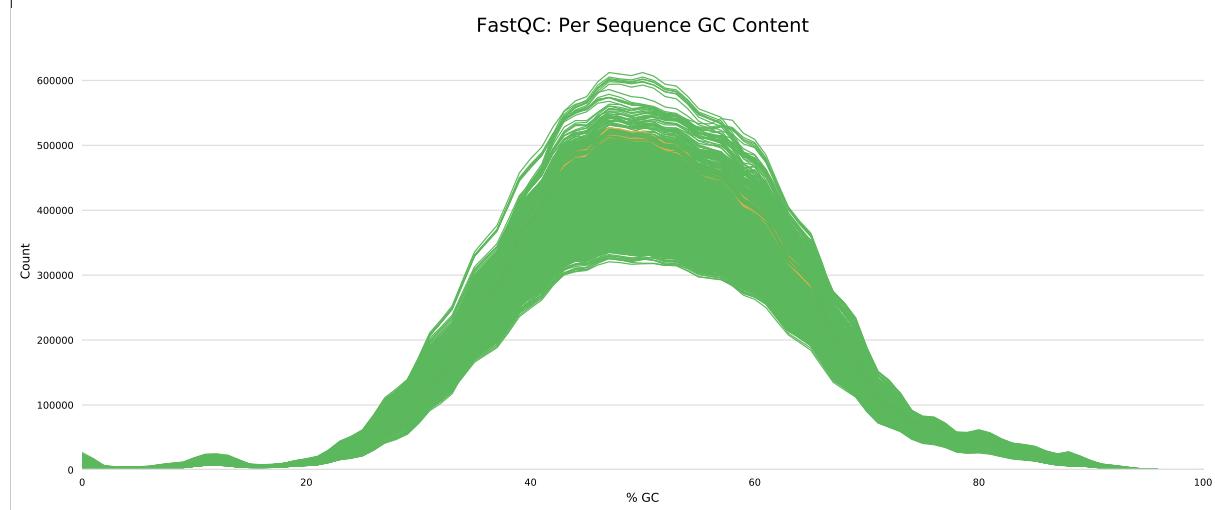


Figure 2.7: FastQC per-read GC distributions for RNA-seq samples. Visualised with MultiQC [191].

[195]) and mapping rate expectations (e.g. 70–90 % [196]).

Genes with short non-coding RNA (ncRNA) biotypes* were filtered out. These are generally not polyadenylated, depleted by size selection during library preparation, and shorter than the 75 bp read length, so expression estimates for these genes can reflect misassignment of counts from overlapping protein-coding or long ncRNA genes [197]. Globin genes, which are highly expressed in RBCs and reticulocytes—cell types expected to be depleted in PBMC [198]—were also filtered out. Given the proportion of removed counts at this stage was low for most samples (Fig. 2.8), poly(A) selection and PBMC isolation were deemed to have been efficient.

Many of the genes in the reference transcriptome were not detectably expressed in PBMC (Fig. 2.9), and many genes were expressed at counts too low for statistical analysis of DGE. Genes were filtered to require a minimum of 0.5 counts per million (CPM) in at least 20 % of samples. The 0.5 CPM threshold was chosen to correspond to approximately 10 counts in the smallest library, where 10–15 counts is a rule of thumb for considering a gene to be robustly expressed [199, 200]. Genes were further filtered to require detection (non-zero expression) in at least 95 % of samples to lessen the impact of low-expression outliers. The change in the distributions of gene expression among samples before and after filtering shows a substantial number of low expression genes are removed (Fig. 2.10).

RNA-seq produces compositional data due to sequencing a fixed number of reads per library; if one gene's expression goes up in a library, another's must go down. In order for expression values to be comparable between different libraries (samples), it is important to account for composition bias: the dependence of expression estimates on the expression properties of other genes in each library [201]. Effective library sizes were computed as between-sample normalisation factors using the trimmed mean of M-values (TMM) method [201, 202] from `edgeR:::calcNormFactors` [203]. Precision weights for each (gene by sample) observation were computed with `limma:::voom` [204] to account for the mean-variance relationship in RNA-seq data; `limma:::voom` also transforms expression values to the \log_2 CPM scale using effective library sizes.

Finally, 15 samples were excluded for having missing HAI or MN data. After the application of all filters, expression values were available for 21 626 genes over 208 samples (70/75 individuals on day 0, 68/75 on day 1, and 70/75 on day 7).

2.2.8 Array data preprocessing

Single-channel Agilent 4x44K expression array data (G4112F, 60-mer oligonucleotide probes) for 173 samples from Sobolev *et al.* [162] were downloaded from ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2313/>). These arrays were originally processed in two batches, the effect of which can be seen in the raw foreground intensities (Fig. 2.11).

`VSN::normalizeVSN` [205] was used for simultaneous background correction, between-array normalisation (affine transformation, centers and scales each array to control for systematic experimental factors), and variance-stabilisation of intensity values (generalised logarithm, similar to \log_2 with better performance for small values), resulting in expression values on a \log_2 scale. As systematic experimental factors might differ between batches, requiring different centering

*miRNA, miRNA_pseudogene, miscRNA, miscRNA pseudogene, Mt rRNA, Mt tRNA, rRNA, scRNA, snlRNA, snoRNA, snRNA, tRNA, tRNA_pseudogene. List from <https://www.ensembl.org/Help/Faq?id=468>

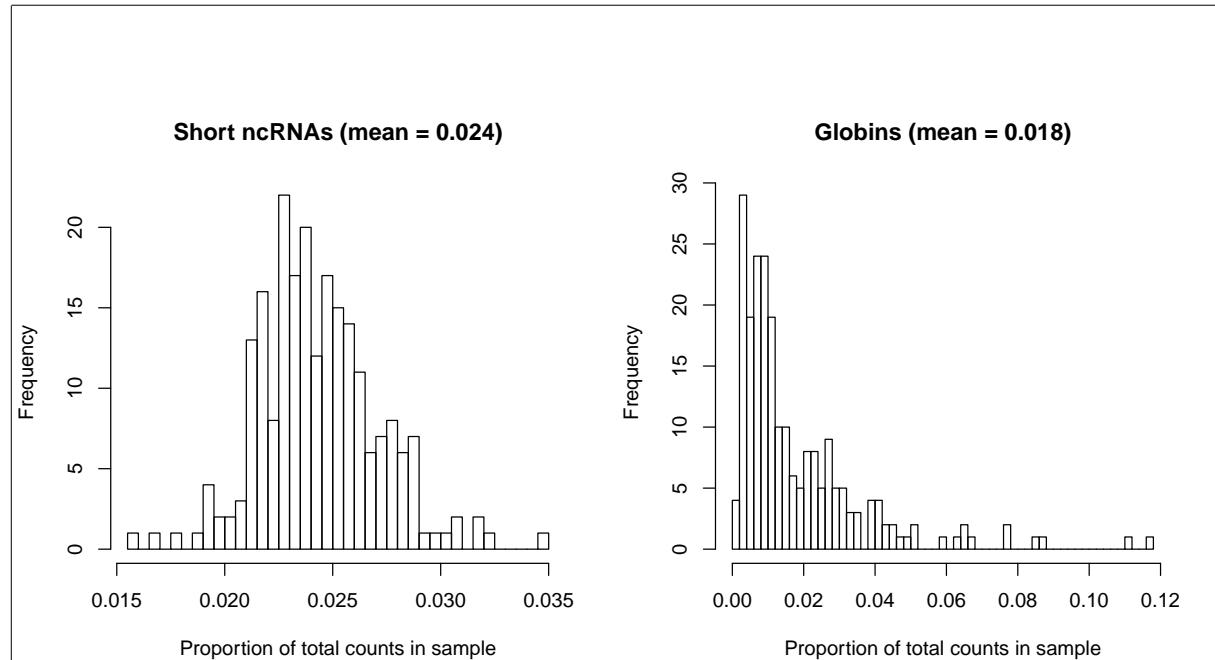


Figure 2.8: Distributions of removed short ncRNA and globin counts as a proportion of total counts in RNA-seq samples.

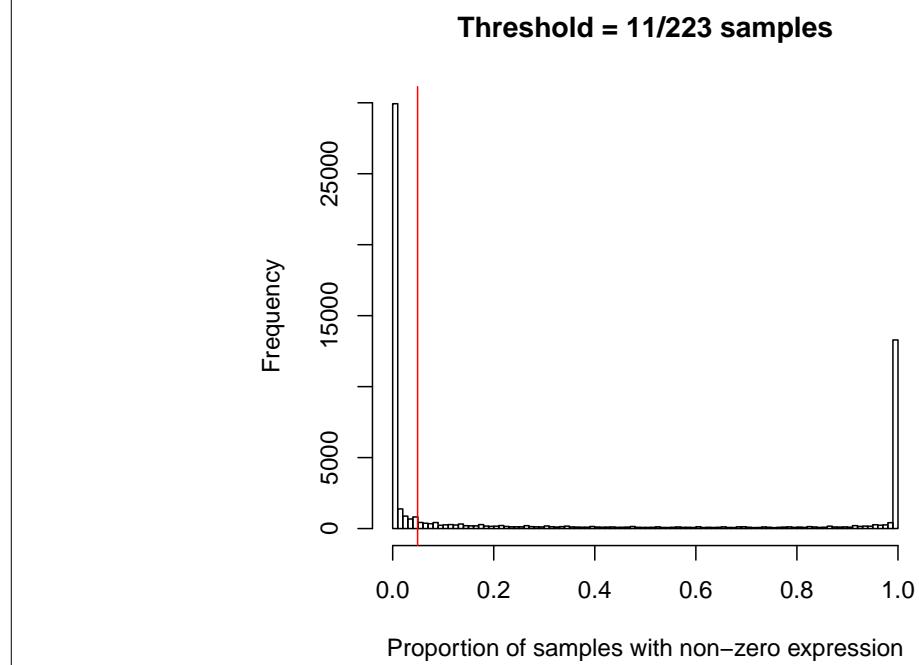


Figure 2.9: Distribution of the proportion of samples in which genes were detected (non-zero expression). Many genes are not detected in any samples (left-hand side). Vertical line shows 5 % threshold below which genes were discarded.

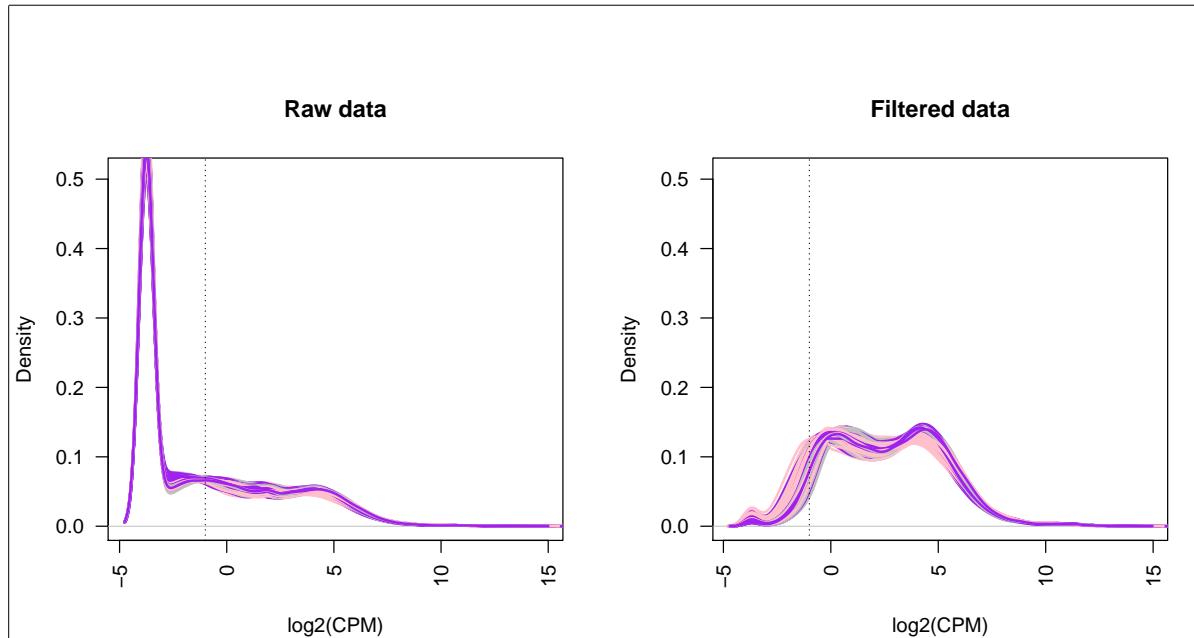


Figure 2.10: Distributions of gene expression for RNA-seq samples before and after filtering low expression and non-detected genes. Vertical line shows $\text{CPM} = 0.5$ threshold.

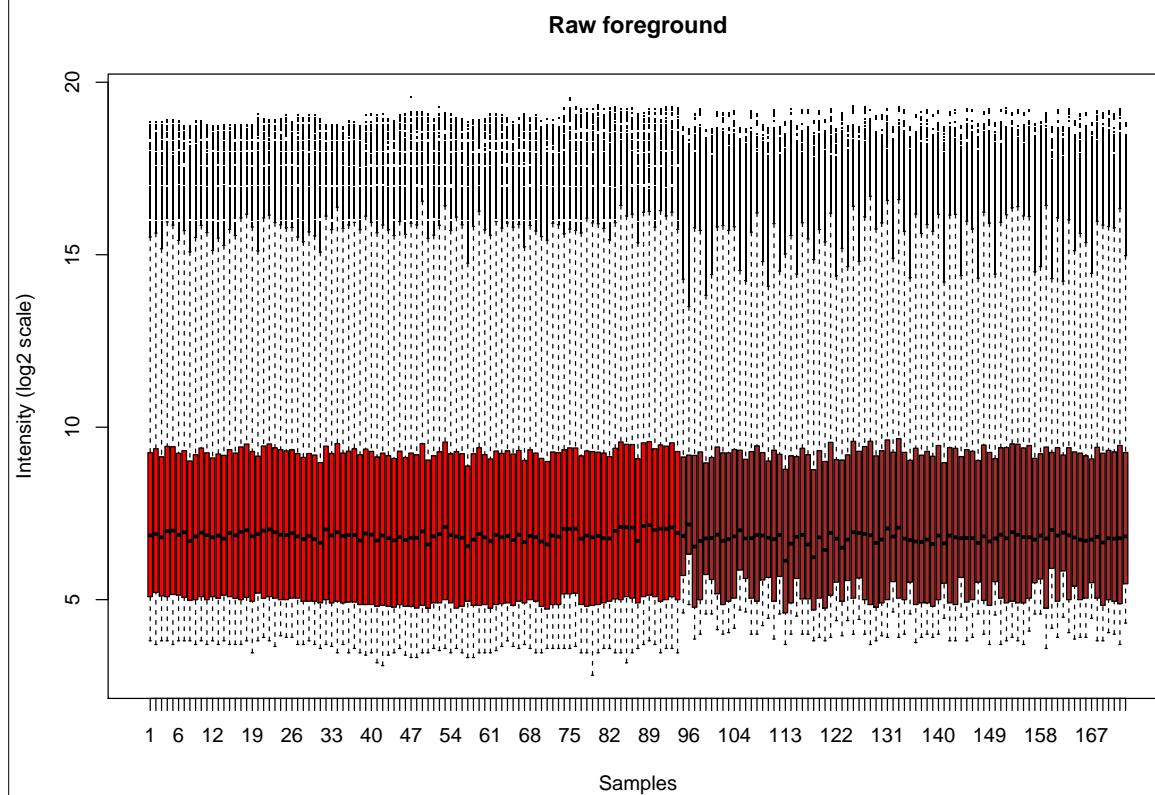


Figure 2.11: Distribution of raw foreground intensities for HIRD array samples ($n = 173$). Colored by array processing batch.

and scaling factors, normalisation was performed per-batch, then the two batches were merged.

Probes were matched to genes using `hgug4112a.db`*. Most genes were targeted by multiple array probes; 31 208 probes were collapsed into 18 216 Ensembl genes using by selecting the probe with the highest mean intensity for each gene (`WGCNA::collapseRows(method = "MaxMean")`, recommended for probe to gene collapsing by Miller *et al.* [206]). While it would be optimal to select a collapsing method to maximise the concordance between array and RNA-seq expression values, there were no samples assayed by both platforms in the **HIRD** dataset. The final normalised \log_2 intensity values for these 18 216 genes over 173 samples is shown in Fig. 2.12. Finally, `limma::arrayWeightsQuick` [207] was used to compute per-sample quality weights used to downweight unreliable arrays (samples) in the **DGE** analyses.

2.2.9 Differential gene expression (DGE)

2.2.9.1 Platform and batch effects

Combining the normalised array and RNA-seq data resulted in expression values for 13 593 genes assayed in both platforms for a total of 374 samples. PCA revealed that although samples separate by experimental timepoint along PC3 (Fig. 2.13e), measurement platform is by far the largest source of variation (Fig. 2.13a). Normalisation was also not able to completely remove the batch effect within the array data (Fig. 2.13a). The large platform effect likely stems from systematic technological differences in how each platform measures expression. RNA-seq has a higher dynamic range, resulting less bias at low expression levels, but estimates are more sensitive to changes in depth than array estimates are to changes in intensity [208]. Agreement between the two platforms is poor at extremes of expression [209, 210]. The preprocessing steps for the two platforms (Sections 2.2.7 and 2.2.8) were also vastly different.

Despite the potential shortcomings of array data detailed above, the array dataset contains individuals with more extreme antibody response phenotypes (Fig. 2.3), and hence should not be excluded. Given the magnitude of the platform effect, I concluded that the appropriate approach was a two-stage approach that meta-analyses per-platform DGE effect estimates while explicitly accounting for between-platform heterogeneity.

Regarding the batch effect within the array data, a popular adjustment method is ComBat [211], which estimates per-gene, per-batch centering and scaling parameters, which are shrunk towards the per-batch mean parameters over all genes using empirical Bayes to improve robustness. ComBat was the method used by Sobolev *et al.* [162]. In comparisons of array batch effect adjustment methods, ComBat performed favourably (versus five other adjustment packages) [212] or comparably (versus fitting batch as a fixed or random effect in the linear model, which are centering-only corrections) [213]. However, where batches are unbalanced in terms of sample size [214] or distribution of study groups that have an impact on expression [215], ComBat can overcorrect batch differences or bias estimates of group differences respectively. In our data, sample size and timepoint groups are fairly balanced between the two array batches (Table 2.2). The proportion of responders is not, but response status does not have as prominent an impact on global expression as timepoint (Fig. 2.13). For the DGE analyses in this chapter, I chose to

*<https://bioconductor.org/packages/release/data/annotation/html/hgug4112a.db.html>

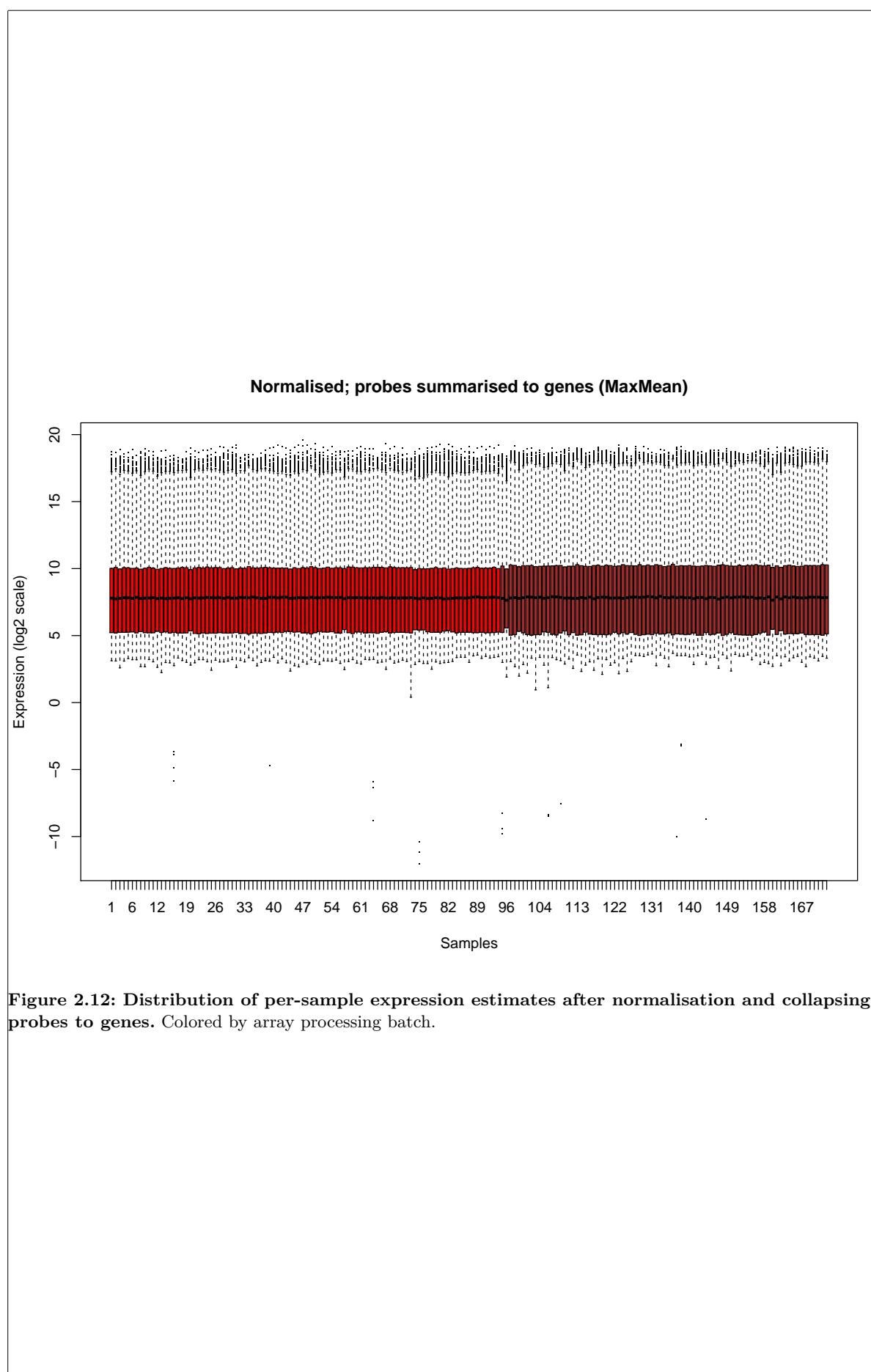


Figure 2.12: Distribution of per-sample expression estimates after normalisation and collapsing of probes to genes. Colored by array processing batch.

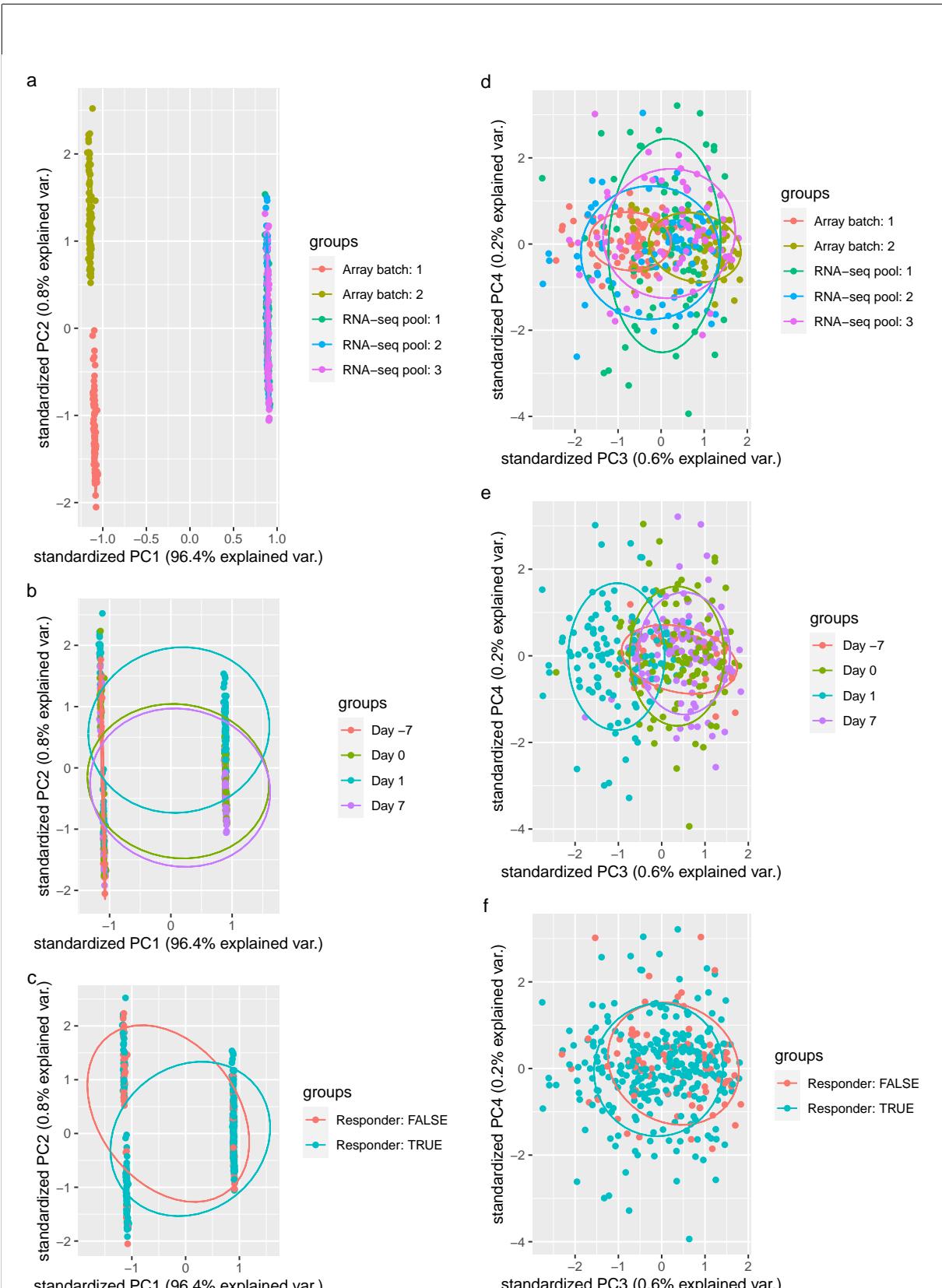


Figure 2.13: First four standardised PCs in the expression data, colored by array batch/RNA-seq pool (a, c), timepoint (b, d), and binary response status (e, f). Expression of each gene was standardised across samples within each platform before PCA.

model batches (array batch and RNA-seq pool) as fixed effects rather than pre-adjusting with ComBat in a separate step, ensuring the degrees of freedom (df) in the DGE model were correct. In practice, results from the analyses were not substantially affected by the choice of whether to use a ComBat pre-adjustment or a fixed effect.

2.2.9.2 Per-platform DGE model

As a meta-analysis was performed, DGE analyses were restricted to the 13 593 genes assayed by both the array and RNA-seq platforms. Linear models were fit using `limma` [216], which is computationally fast, performs well for sufficiently large ($n \geq 3$ per group) sample sizes [217], and internally considers the precision weights computed for RNA-seq observations in Section 2.2.7, and the array quality weights computed for array samples in Section 2.2.8. As Sobolev *et al.* [162] already found there was no global dissimilarity in array expression between day -7 and day 0, for the DGE analyses in this chapter, array day -7 and day 0 are treated as repeated measurements taken at a single “baseline” timepoint.

For each gene and platform, I fit a model (model 1) with expression as the response variable; with an intercept, timepoint (baseline, day 1, day 7), **TRI**, array batch/RNA-seq pool, sex, age, and the first 4 genotype PCs as fixed-effect predictors; and individual as a random-effect predictor. Within-individual correlations for the random effect were estimated using `limma::duplicateCorrelation`. A second model (model 2) was also fit, the only difference being two additional predictors for the multiplicative interactions between day 1 and day 7 with **TRI**. Model 1 was used for testing differences in expression between pairs of timepoints, and for testing association between **TRI** and expression with timepoints pooled. Model 2 was used for testing association between **TRI** and expression at specific timepoints.

Contrasts were defined, testing if linear combinations of estimated model coefficients are different from zero. From model 1, I defined contrasts for day 1 vs. baseline, day 7 vs. baseline, day 7 vs. day 1, **TRI**, sex, and age. For example, to test for association between **TRI** and expression, I used a contrast where the weight for the **TRI** coefficient was 1, with all other coefficient weights set to 0; to test for differences between day 7 vs. day 1, I used a contrast where the weight for the day 7 coefficient was 1, the weight for the day 1 coefficient was -1, and all other coefficient weights were 0. From model 2, I defined contrasts for the **TRI**, **TRI**-day 1, and **TRI**-day 7 interaction terms, which respectively test for association between **TRI** and expression at specifically at baseline, day 1, and day 7. Corresponding coefficients and standard errors for the contrasts were extracted from the `limma` models, which represent effect size in units of \log_2 expression fold change per unit change in predictor value.

2.2.9.3 Choice of DGE meta-analysis method

Two popular frameworks for effect size meta-analysis are fixed-effect and random-effects [218, 219]. The fixed-effect model assumes a single true effect size θ common to all studies. Given k studies ($i = 1, \dots, k$), the observed effect size in the i th study is commonly assumed to be $y_i \sim \mathcal{N}(\theta, \sigma_i^2)$, where observed variation is explained only by within-study sampling error σ_i . In meta-analysis, the effects are combined with some weighting, commonly the inverse variance (precision) $1/\sigma_i^2$.

Table 2.2: Distribution of HIRD samples among timepoint and responder groups in the array batches and RNA-seq pools. Values are count and percentage for categorical variables; mean and standard deviation for continuous variables.

	Total n = 374	Array 1 n = 87	Array 2 n = 79	RNA-seq 1 n = 70	RNA-seq 2 n = 69	RNA-seq 3 n = 69
Day						
-7	40 (10.7%)	20 (23%)	20 (25.3%)	0 (0%)	0 (0%)	0 (0%)
0	114 (30.5%)	24 (27.6%)	20 (25.3%)	24 (34.3%)	23 (33.3%)	23 (33.3%)
1	109 (29.1%)	21 (24.1%)	20 (25.3%)	22 (31.4%)	23 (33.3%)	23 (33.3%)
7	111 (29.7%)	22 (25.3%)	19 (24.1%)	24 (34.3%)	23 (33.3%)	23 (33.3%)
Responder						
FALSE	80 (21.4%)	12 (13.8%)	36 (45.6%)	11 (15.7%)	9 (13%)	12 (17.4%)
TRUE	294 (78.6%)	75 (86.2%)	43 (54.4%)	59 (84.3%)	60 (87%)	57 (82.6%)
TRI						
	-0.1 (1.0)	-0.1 (1.0)	-0.4 (1.4)	0.1 (0.6)	-0.0 (0.8)	0.2 (0.6)

The random-effects model assumes a distribution of true effects centered around a common mean μ . Each of the k studies estimates its own study-specific true effect size θ_i . These are distributed around μ with variance τ^2 (standard deviation τ), representing an additional source of variation: the between-study heterogeneity. Then we have $y_i \sim \mathcal{N}(\theta_i, \sigma_i^2)$ for the first level of variation, $\theta_i \sim \mathcal{N}(\mu, \tau^2)$ for the second level of variation, and assuming these distributions, we have a normal-normal multilevel model [220]. Study weights include both within- and between-study variance $1/(\sigma_i^2 + \tau^2)$, reducing to the fixed-effect model when $\tau = 0$.

The choice of fixed or random effects depends on whether it is tenable to assume studies are identical enough that they all estimate a common effect*. In the HIRD data, there are $k = 2$ studies: array and RNA-seq. The between-platform differences described in Section 2.2.9.1 represent considerable sources of between-study heterogeneity. For DGE effect sizes, arrays also suffer from ratio compression of fold change estimates due to cross-hybridisation and probe saturation [210, 222, 223]. The assumption of $\tau = 0$ is unrealistic, so a random-effects model is more appropriate.

Unfortunately, there is no optimal solution for directly estimating τ in random-effects meta-analyses with small k [224], and especially in the case of $k = 2$ [225]. Many estimators are available [226], but lack of information with small k causes estimation to be imprecise, and often results in boundary values of $\tau = 0$ that are incompatible with the assumed positive heterogeneity [227, 228]. In such circumstances, the most sensible approach may be to incorporate prior information about hyperparameters μ and τ in a Bayesian random-effects framework [226–229]. For this study, I used the implementation in `bayesmeta` [220].

*A common misinterpretation is that random-effects meta-analysis assumes studies *themselves* are sampled from a population of studies. This is rarely appropriate since the design of new studies is influenced by existing studies [221]. The required assumption is exchangeability of study *effects*, which informally states effects are neither completely identical nor completely independent, but “similar” [221].

2.2.9.4 Prior for between-study heterogeneity

The choice of prior for between-study heterogeneity τ is influential when k is small [229]. Gelman [230] considers the case of $k = 3$, showing that a flat prior places too much weight on implausibly large estimates of τ , and recommends a weakly informative prior that acts to regularise the posterior distribution, constraining it away from implausible values. Since I assumed zero estimates for τ are unrealistic, I used a weakly informative gamma prior, as recommended by Chung *et al.* [227], which has zero density at $\tau = 0$ but increases gently as τ increases (a positive constant derivative at zero). This constrains τ to be positive, but still permits estimates close to zero if the data support it. This is in contrast to priors used in other studies from the log-normal (e.g. [231, 232]) or inverse-gamma (e.g. [233]) families that have zero density at zero and derivatives of zero close to zero, ruling out small values of τ no matter what the data suggest; and in contrast to half- t family priors (e.g. [229, 230]), which have their mode at zero, and do not rule out $\tau = 0$.

Instead of constraining the value of τ for a gene's effect size to be a single unreliable estimate from $k = 2$ data points, assuming a prior distribution recognises that other genes may be informative of the range of plausible values for between-platform heterogeneity. To estimate the appropriate shape and scale parameters for the gamma empirically, a frequentist random-effects model using the **restricted maximum likelihood (REML)** estimator for τ (recommended for continuous effects [226]) was fit for each gene using `metafor::rma.uni` [234]. Depending on the contrast, over half of resulting per-gene τ estimates were boundary values of zero. Small estimates of $\tau < 0.01$ were excluded, and a gamma distribution fit to the remaining estimates using `fitdistrplus` [235].

2.2.9.5 Prior for effect size

While the choice of prior on τ is influential when k is small, there is usually enough data to estimate the effect size μ such that any reasonable non-informative prior can be used [228, 230]. `bayesmeta` implements both flat and normal priors for μ . Assuming that most genes are not differentially expressed with effect sizes distributed randomly around zero, I selected a normal prior with $N(\mu = 0, \sigma^2)$, over a flat prior. As in the section above, to determine an appropriate scale, a normal distribution with mean $\mu = 0$ was fit to the distribution of effect sizes from the per-gene frequentist models to empirically estimate σ .

Heavy-tailed Cauchy priors have been proposed for effect size distributions in DGE experiments to avoid over-shrinkage of true large effects in the tails [236]. Since `bayesmeta` does not implement a Cauchy prior, to avoid over-shrinkage, I flatten the normal prior considerably by scaling up the standard deviation by a factor of 10: $N(0, (10\sigma)^2)$. This places a 95 % prior probability that effects are less extreme than approximately 20 times the observed σ , sufficient to allow for extreme fold-changes.

2.2.9.6 Example of priors

An example of the empirically estimated hyperparameters for the priors for the day 1 vs. baseline contrast are shown in Fig. 2.14 (for τ) and Fig. 2.15 (for μ). For τ , the final prior used was `Gamma(shape = 1.57, scale = 0.06)`. This is comparable to the default recommendation from

Chung *et al.* [227] of a $\text{Gamma}(\text{shape} = 2, \text{scale} = \lambda)$ prior where λ is small. For μ , the final prior used was $N(0, (0.324 \times 10)^2)$. The tails of the non-scaled normal fit (black) are light compared to the Cauchy fit (red), which may lead to over-shrinkage, especially since there are many genes with high positive fold changes for the day 1 vs. baseline effect.

2.2.9.7 Multiple testing correction

For per-platform DGE, false discovery rate (FDR) was controlled with `limma::decideTests` using the Benjamini-Hochberg (BH) procedure. For the frequentist random-effects meta-analysis, nominal per-gene p -values were converted to FDR estimates using `p.adjust(method = "BH")` in R. For the Bayesian random-effects meta-analysis, the effect sizes and standard errors from the per-gene meta-analysis output from `bayesmeta` were supplied to `ashr` [238], which models the distribution of effects under the assumption of unimodality. `ashr` applies empirical Bayes shrinkage to improve the accuracy of effect estimation (e.g. against winner's curse), returning posterior effect sizes, posterior standard errors, and their significance (local false sign rate (LFSR)). LFSR is analogous to FDR, but quantifies the probability, given the data, of calling the wrong sign for an effect, rather than the confidence of a non-zero effect [238]. Unless otherwise stated, FDR and LFSR were controlled at the 5 % level separately for each contrast, as control is for the proportion of positives expected to be false positives, which is scalable to multiple contrasts.

2.2.10 Ranked gene set enrichment using blood transcription modules

The gene sets used were blood transcription modules (BTMs) from Chaussabel *et al.* [239] (prefixed “DC”) and Li *et al.* [240] (prefixed “LI”). Modules are sets of genes with transcriptional and functional similarities across a variety of healthy, diseased, and stimulated conditions. The 260 modules from Chaussabel *et al.* [239] were constructed by unsupervised clustering of 239 PBMC transcriptomes from multiple disease datasets, then annotated by data mining of gene names in PubMed abstracts. The 334 modules from Li *et al.* [240] were constructed from coexpression analysis of approximately 30 000 blood transcriptomes, then annotated making use of Gene Ontology (GO) terms, cell type-specific markers, pathway databases, and manual literature searches. These datasets are particularly suitable for systems vaccinology studies, given their focus on the blood transcriptome. Li *et al.* [240] modules are better annotated in general, and were used for the majority of gene set enrichments in this chapter.

Gene set enrichment analyses were conducted using `tmod::tmodCERNOtest` [241], which assesses the enrichment of small ranks within specific sets of genes compared to all genes, after the genes are ranked by some metric—here I used effect sizes from `bayesmeta`. The CERNO statistic for a gene set is:

$$-2 \sum_{i=1}^n \ln \frac{r_i}{N} \sim \chi^2(2n) \quad (2.1)$$

where n is the number of genes in the set, N is the number of measured genes in the experiment, and $r_i \in 1, 2, \dots, N$ is the rank of the i th gene in the set among all measured genes. CERNO is relatively robust to the ranking metric [242]. FDR control for the number of gene sets tested was performed using BH, again separately for each contrast. The χ^2 test is one-sided, so `tmod::tmodCERNOtest` only considers enrichment of small ranks when computing significance.

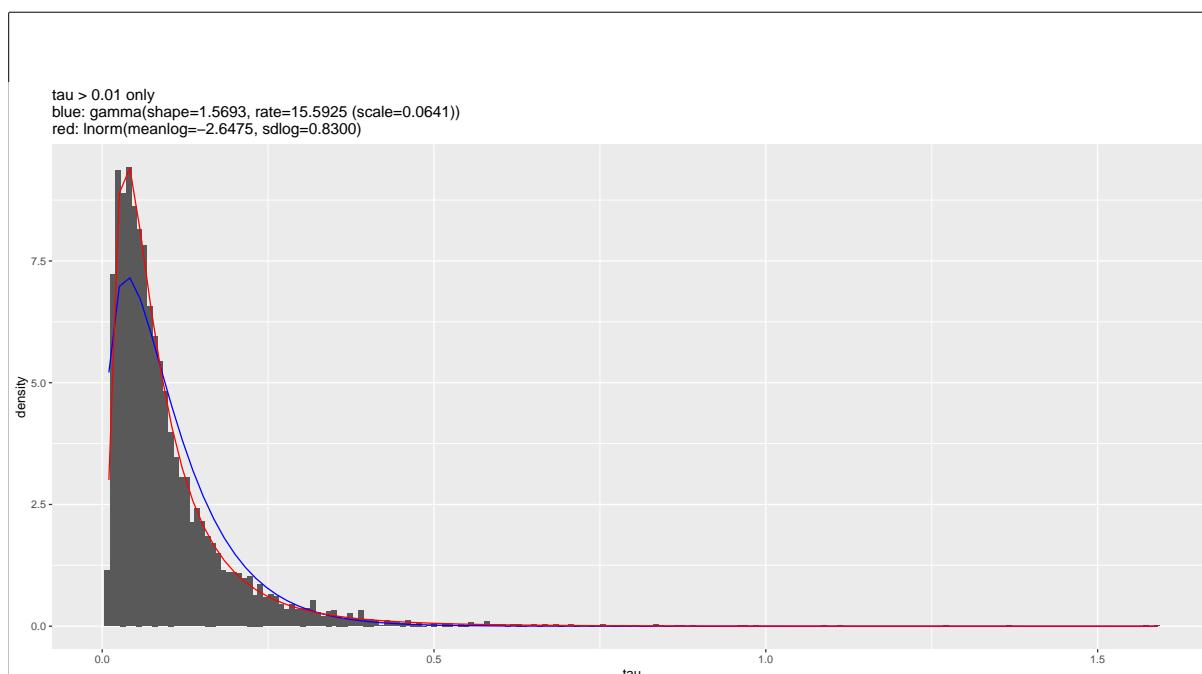


Figure 2.14: Gamma prior for τ (blue) used for `bayesmeta` analyses of the day 1 versus baseline effect, compared to the empirical distribution of per-gene frequentist `metafor::rma.uni` estimates for τ . Genes with small estimates of $\tau < 0.01$ were excluded before distribution fitting. Empirical log-normal fit also shown (red). Distribution parameters are listed.

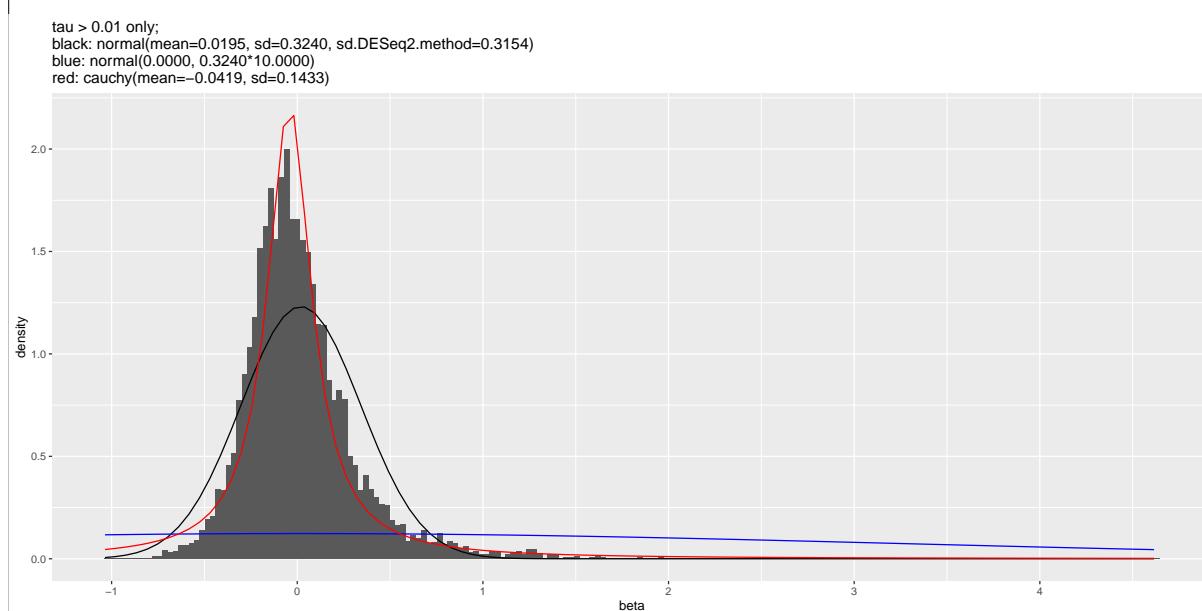


Figure 2.15: Normal prior for μ (blue) used for `bayesmeta` analyses of the day 1 versus baseline DGE effect, compared to the empirical distribution of per-gene frequentist `metafor::rma.uni` estimates for τ . Genes with small estimates of $\tau < 0.01$ were excluded before distribution fitting. The original non-scaled normal fit is shown (black), as well as a Cauchy fit (red). Distribution parameters are listed. An alternative estimate of the Normal standard deviation more robust to outliers using a quantile matching method from `DESeq2` [237] is also given. In this case, it was comparable to the maximum likelihood (ML) estimate from `fitdistrplus`.

As genes can be down or upregulated, separate tests were performed sorting genes in ascending and descending order, and the more significant result was used to determine the overall direction of effect for each gene set. As the approach is rank-based and considers all measured genes, no filters based on the ranking metric were necessary.

The effect size of a gene set enrichment can be quantified with the **area under the curve (AUC)**, computed from U , the test statistic from a Mann-Whitney U test (also known as the Wilcoxon rank-sum test):

$$U = n(N - n) + \frac{n(n + 1)}{2} - \sum_{i=1}^n r_i \quad (2.2)$$

This is a non-parametric test for whether genes in the set have smaller ranks than genes not in the set on average. Then $\text{AUC} = U/(n(N - n))$, which takes values from 0 to 1. Significant results from the one-sided `tmod::tmodCERN0test` will always have $\text{AUC} > 0.5$.

2.3 Results

2.3.1 Extensive global changes in expression after vaccination

To gain an overview of how the transcriptome changes after vaccination, linear models were fit to identify genes differentially expressed at day 1 or day 7 compared to baseline (day -7 and day 0) in the **HIRD** array and **RNA-seq** expression data, accounting for covariates such as batch effects, sex, age, **TRI**, and ancestry. At the 13 593 genes with expression measured by both platforms, models were fit within each platform. A frequentist random-effects meta-analysis was initially run to generate plausible values for **DGE** effect size and between-platform heterogeneity. These were used to form empirical priors for a Bayesian random-effects meta-analysis, producing final posterior estimates of effect size and standard errors.

Vaccination induced changes in a large proportion of the **PBMC** transcriptome; 6257/13 593 genes were differentially expressed between any pair of timepoints (**LFSR** < 0.05). Applying an absolute **FC** > 1.5 cutoff identified 857 genes with the strongest effects. Their expression clustered into three general patterns: upregulation from baseline to day 1, then downregulation from day 1 to day 7 back to baseline levels; upregulation from baseline to day 1, sustained at day 7; and downregulation from baseline to day 1, then upregulation from day 1 to day 7 back to baseline levels (Fig. 2.16).

2.3.1.1 Innate immune response at day 1 post-vaccination

Consistent with global expression at day 1 being markedly different from expression at other timepoints (Fig. 2.13), the highest numbers of differentially expressed genes were observed at day 1, with 644 genes differentially expressed vs. baseline. The majority of these (580/644) were upregulated. The gene with the highest **FC** increase at day 1 compared to baseline was *ANKRD22* ($\log_2 \text{FC} = 4.49$), an interferon-induced gene in monocytes and **DCs** involved in antiviral innate immune pathways [243]. Other key genes in the interferon signalling pathway [244] such as *STAT1* ($\log_2 \text{FC} = 2.17$), *STAT2* ($\log_2 \text{FC} = 0.95$), and *IRF9* ($\log_2 \text{FC} = 0.82$) were also upregulated at day 1. Rank-based gene set enrichment analysis using `tmod` [241] revealed that genes with

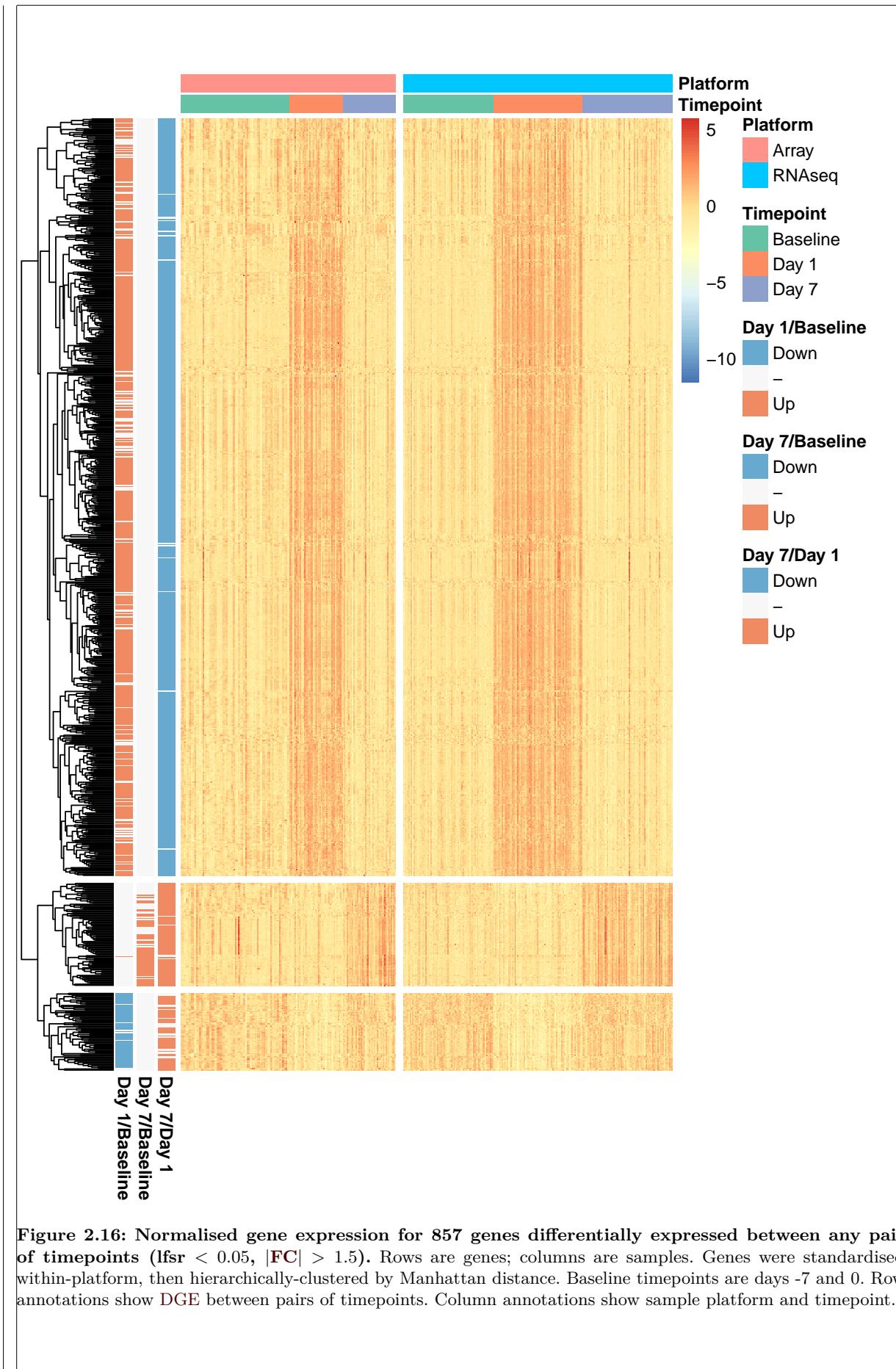


Figure 2.16: Normalised gene expression for 857 genes differentially expressed between any pair of timepoints ($\text{lfsr} < 0.05$, $|\text{FC}| > 1.5$). Rows are genes; columns are samples. Genes were standardised within-platform, then hierarchically-clustered by Manhattan distance. Baseline timepoints are days -7 and 0. Row annotations show DGE between pairs of timepoints. Column annotations show sample platform and timepoint.

the large FC increases at day 1 were enriched in modules associated with interferon, activated DCs, monocytes, and toll-like receptors (TLRs) and inflammatory signalling (Fig. 2.17). Sobolev *et al.* [162] reported only a 1.6-fold ($\log_2 1.6 = 0.68$) increase in blood monocytes from baseline to day 1, as measured by FACS, so these changes reflect active, per-cell upregulation as well as proliferation.

Sixty-four genes were downregulated at day 1, enriched in modules associated with T cells and natural killer (NK) cells. The largest absolute fold change was observed for *FGFBP2* ($\log_2 \text{FC} = -0.91$), which encodes Ksp37, a secretory protein unique to CD8⁺ T cells and NK cells [245]. Again, the fold changes in expression were of greater magnitude than observed for the abundance of these cell types, suggesting active downregulation Sobolev *et al.* [162].

As can be seen in Fig. 2.16, there was a general tendency for expression to return to baseline levels by day 7. This was the case for 566/644 upregulated genes and 44/64 downregulated genes, indicating the innate phase of response likely peaks in the first few days.

2.3.1.2 Adaptive immune response at day 7 post-vaccination

Fifty-nine genes were differentially expressed at day 7 vs. baseline. The genes with the highest upregulation were genes associated with B cell differentiation and maturation: *TNFRSF17* (marginal zone B and B1 cell specific protein, $\log_2 \text{FC} = 1.75$) and *MZB1* (B-cell maturation antigen, $\log_2 \text{FC} = 1.74$). Genes specific to plasma cells, including *SDC1* (which encodes CD138, required for plasma cell maturation [246]) ($\log_2 \text{FC} = 1.37$) and *ELL2* (which mediates antibody secretion [247]) ($\log_2 \text{FC} = 0.87$) were also prominently upregulated. This matches an almost 5-fold increase in plasma cell abundance at day 7 compared to baseline [162]. Strongly enriched modules at day 7 were related to mitosis and cell proliferation, particularly in CD4⁺ T cells (Fig. 2.17). Both the CD4⁺ T cell and plasma cell response are indications of a shift toward an adaptive and primarily humoral immune response by day 7.

2.3.2 Expression associations with antibody response

2.3.2.1 Between-platform heterogeneity hinders detection of gene-level associations

Using only array expression data, Sobolev *et al.* [162] identified a set of 62 genes with day 7 expression associated with antibody response, where response was defined as a binary phenotype based on 4-fold increases in HAI or MN titres from day -7 to day 63. Many of these genes were related to plasma cell development and antibody production. I aimed to find genes similarly associated with antibody response in the meta-analysis of array and RNA-seq expression data, and assess the replication of the 58/62 genes that fell into the set of 13 593 genes measured by both platforms.

I computed a baseline-adjusted, continuous measure of antibody response, the TRI [153]. The TRI is comparable to the binary definition in ranking (Fig. 2.2g, Fig. 2.2h), but as a continuous phenotype, it improves statistical efficiency to detect associations. Within just the array data, 51/58 genes were replicated (FDR < 0.05), confirming TRI and the binary response phenotype were comparable. However, using only the RNA-seq data replicated 0/58 genes.

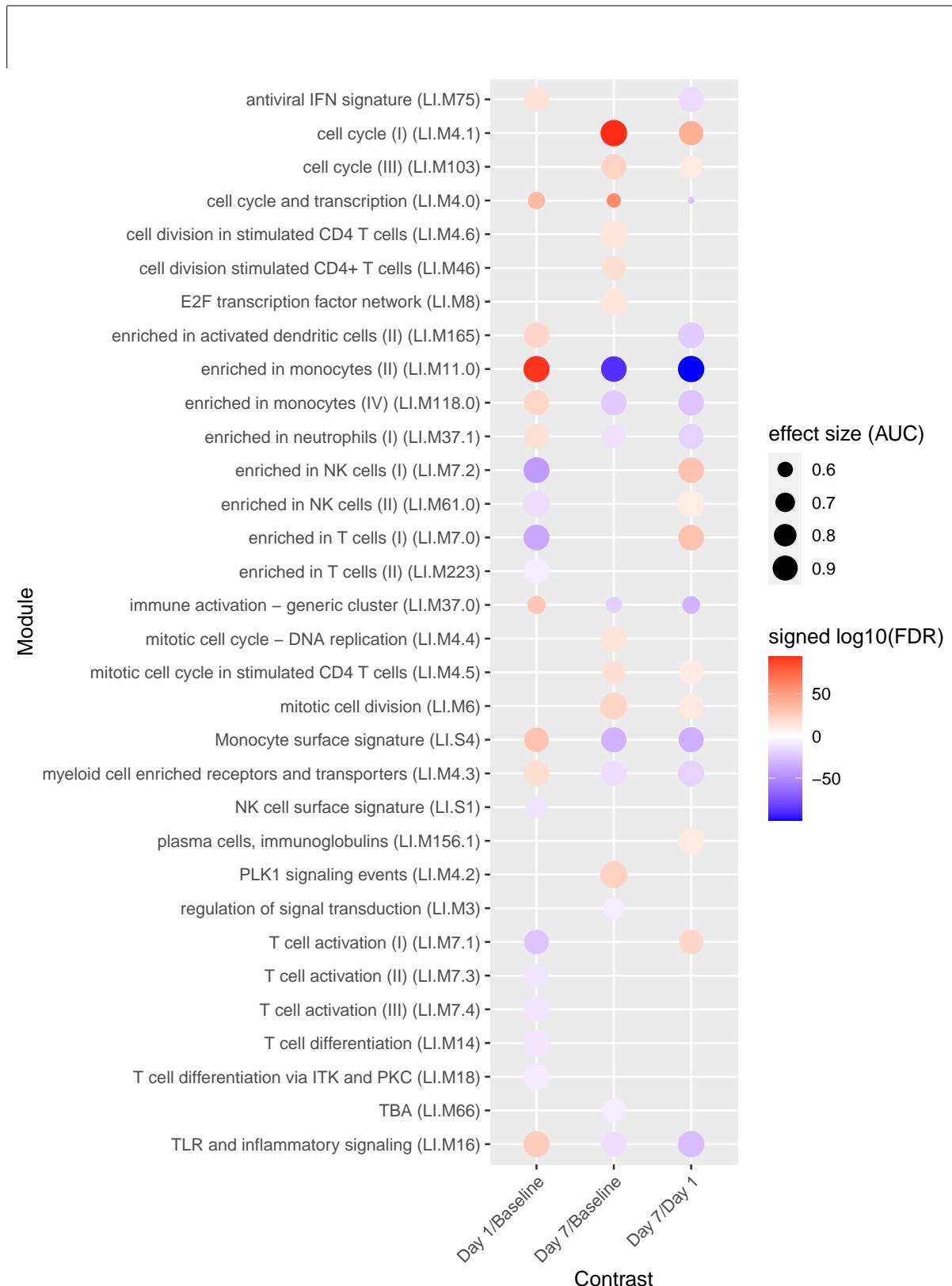


Figure 2.17: Transcriptomic modules up or downregulated between pairs of timepoints. The top ten most significant modules for each contrast are shown. Size of circle indicates absolute effect size (AUC). Color of circle indicates significance ($FDR < 0.05$) and direction of effect (red = upregulation, blue = downregulation). Absence of circle indicates non-significance.

In the initial frequentist random-effects meta-analysis, with a significance threshold of $\text{FDR} < 0.05$, 6 genes had expression associated with **TRI** at baseline (Fig. 2.18f), 55 at day 7 (Fig. 2.18h), and 11 pooling samples over all timepoints (Fig. 2.18e). Of the day 7-specific associations reported by Sobolev *et al.* [162] (circled in Fig. 2.18h), 15/58 replicated, all with the same positive direction of effect (high expression with high **TRI**). However, almost all significant results displayed higher effect sizes in the array compared to **RNA-seq** (13/15 genes). This was in contrast to the associations identified with timepoint, where significant genes had more consistent effects between platforms along the diagonal (Fig. 2.18b–d). The likely cause is the presence of more extreme antibody response phenotypes (higher **TRI** range) in the array versus the **RNA-seq** dataset (Fig. 2.3). This represents an additional source of between-platform variation not due to technical factors, but inherent to the samples themselves.

The Bayesian meta-analysis pipeline more robustly models between-study heterogeneity due to platform and sample-specific effects. Due to shrinkage of effects, few genes with effects closer to the dense center of the effect distribution were called as significant, and significant genes tended to have outlying effect sizes in both platforms (compare Fig. 2.18b–d with Fig. 2.19b–d). No single gene was detected as significantly associated with **TRI** at $\text{LFSR} < 0.05$ for any contrast; not at any single timepoint, nor when pooling samples across all timepoints (Fig. 2.19e–h). The frequentist meta-analysis is likely to use poor estimates of the between-platform heterogeneity, as there are only two data points to estimate it from. Indeed, all 15 significant genes with day 7 expression associated with **TRI** in the frequentist meta-analysis had unrealistic between-platform heterogeneity estimates of exactly zero (Fig. 2.20).

2.3.2.2 Module-level associations with antibody response

Using effect sizes from the Bayesian meta-analysis, significant enrichments were detectable at the gene set level. The strongest effects were seen at day 7, where expression of modules related to the cell cycle, CD4⁺ T cells, and plasma cells were positively associated with **TRI**—“cell cycle (I)” (LI.M4.1, $\text{FDR} = 6.81 \times 10^{-54}$), “Plasma cell surface signature” (LI.S3, $\text{FDR} = 1.78 \times 10^{-12}$), and “cell division stimulated CD4⁺ T cells” (LI.M46, $\text{FDR} = 5.54 \times 10^{-10}$) (Fig. 2.21).

Associations with **TRI** were also detected at baseline. A diverse set of set of modules had positive associations, including “chemokines and inflammatory molecules in myeloid cells” (LI.M86.0, $\text{FDR} = 2.25 \times 10^{-11}$), “platelet activation - actin binding” (LI.M196, $\text{FDR} = 1.71 \times 10^{-8}$), “enriched in B cells (I)” (LI.M47.0, $\text{FDR} = 2.40 \times 10^{-7}$), “cell adhesion” (LI.M51, $\text{FDR} = 1.22 \times 10^{-10}$), “myeloid, dendritic cell activation via NFkB (I)” (LI.M43.0, $\text{FDR} = 4.68 \times 10^{-7}$), and “proinflammatory dendritic cell, myeloid cell response” (LI.M86.1, $\text{FDR} = 4.11 \times 10^{-7}$). Monocyte modules “enriched in monocytes (II)” (LI.M11.0, $\text{FDR} = 3.53 \times 10^{-4}$) and “Monocyte surface signature” (LI.S4, $\text{FDR} = 1.17 \times 10^{-3}$) were negatively association with **TRI**. Negative associations for these same modules were also maintained at day 1 (LI.M11.0, $\text{FDR} = 1.41 \times 10^{-10}$; LI.S4, $\text{FDR} = 1.74 \times 10^{-6}$) and at day 7 (LI.M11.0, $\text{FDR} = 5.54 \times 10^{-10}$) (Fig. 2.21).

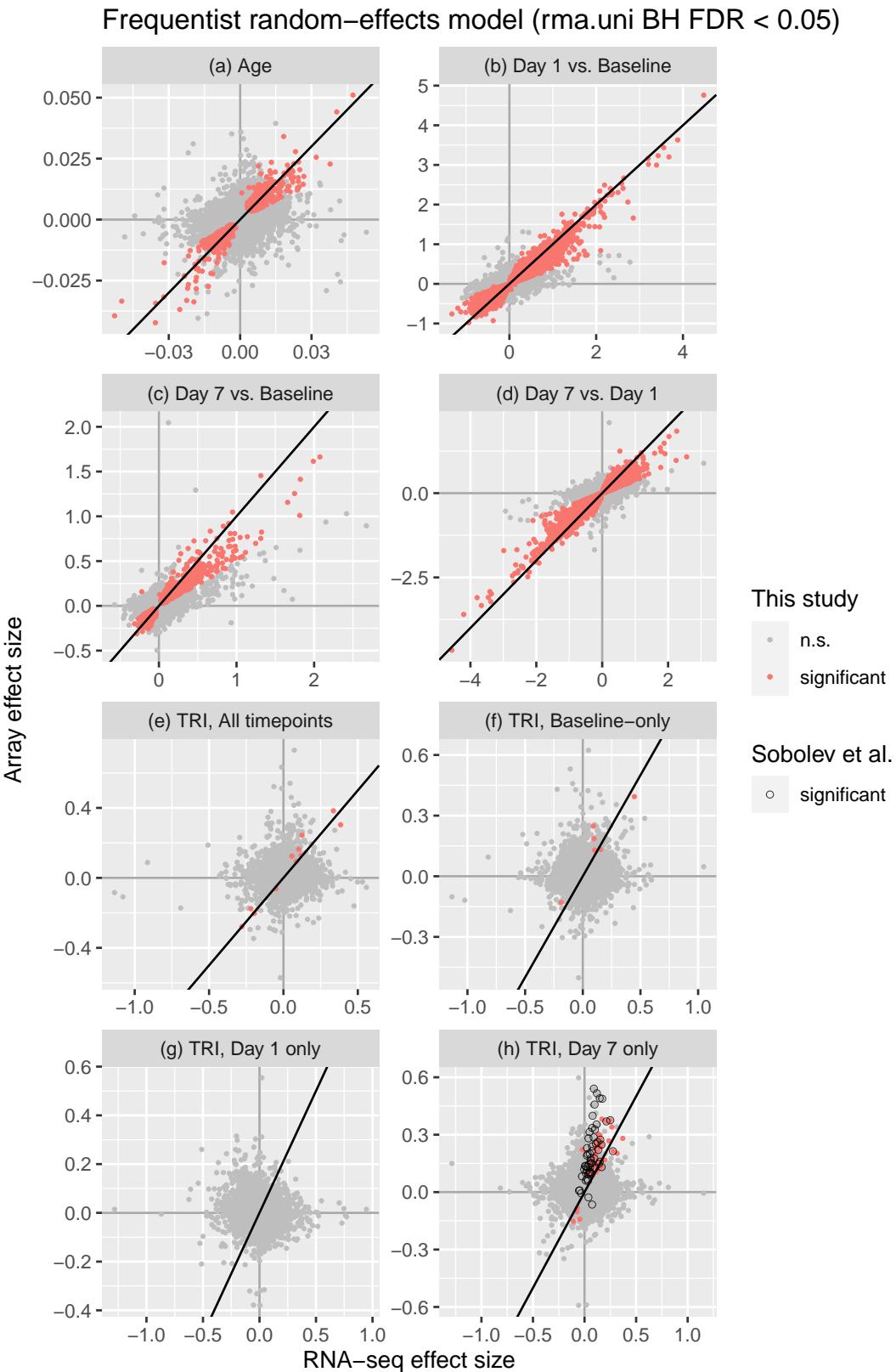
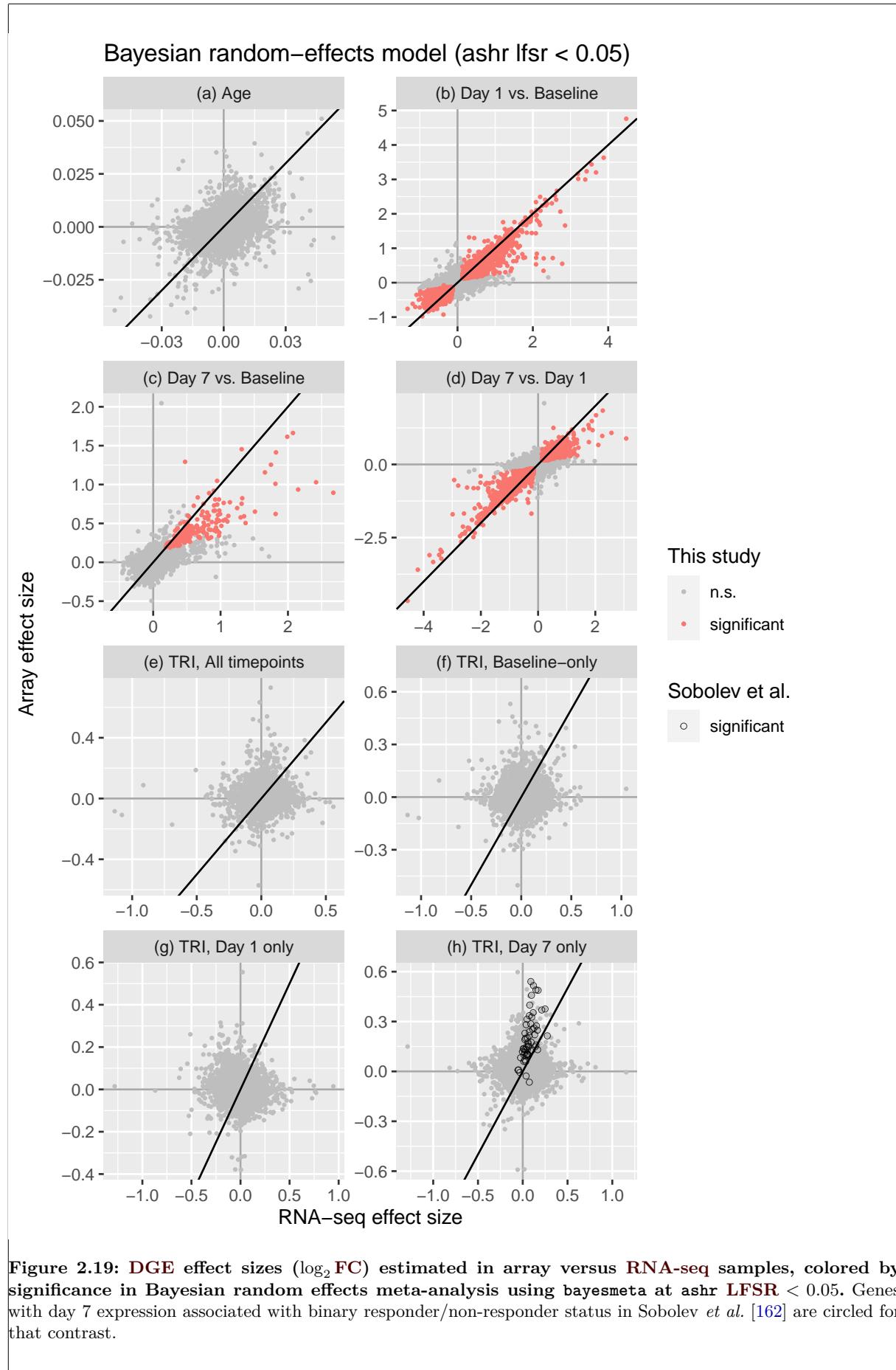


Figure 2.18: DGE effect sizes ($\log_2 \text{FC}$) estimated in array versus RNA-seq samples, colored by significance in frequentist random effects meta-analysis using rma.uni at BH FDR < 0.05. Genes with day 7 expression associated with binary responder/non-responder status in Sobolev *et al.* [162] are circled for that contrast.



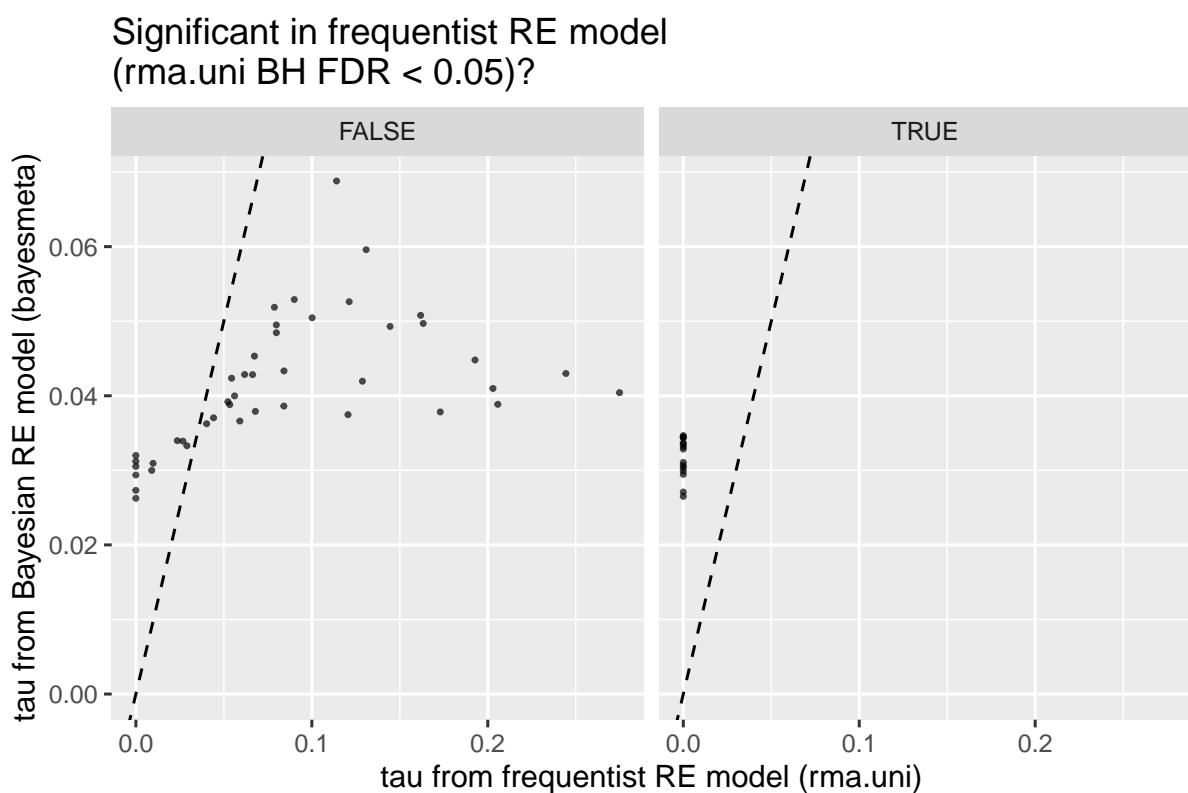


Figure 2.20: Estimates of between-platform heterogeneity τ from frequentist and Bayesian meta-analysis, for the 58 genes with a significant association between day 7 expression and binary responder/non-responder status in Sobolev *et al.* [162]. Dashed line is the identity line. Estimates from the frequentist method cover a wide range and can be zero. For this contrast testing association between day 7 expression and TRI, 8563/13 593 of per-gene τ estimates are zero, including all 15/58 significant results (right). Significant results are array-driven, with 13/15 having higher effects in array than RNA-seq (54/58 genes overall). Estimates of τ from the Bayesian method are in a narrower range and constrained away from zero by the prior.

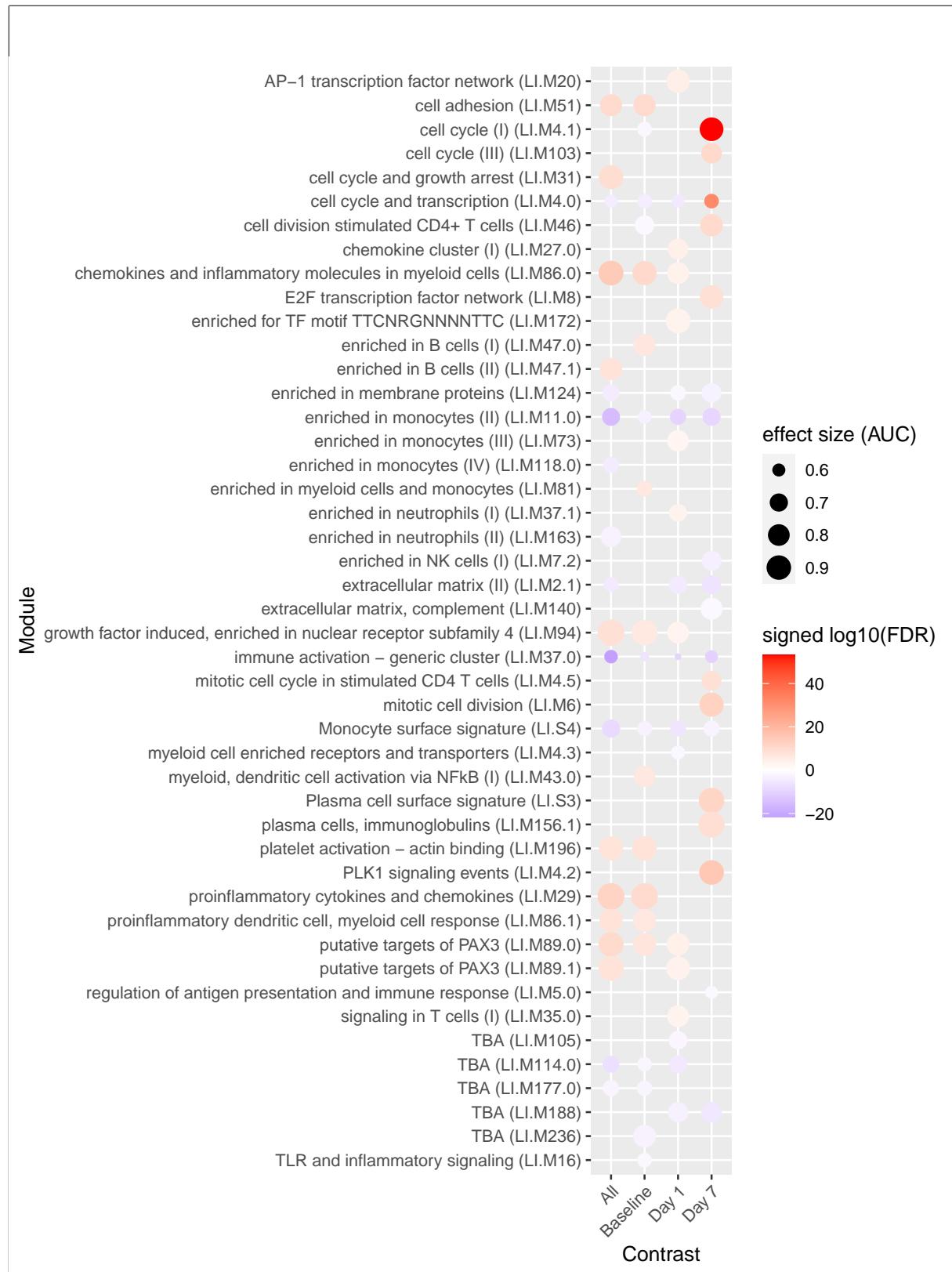


Figure 2.21: Gene expression modules associated with antibody response (TRI). Enrichments were performed with all timepoints pooled, and at each timepoint specifically. The top ten most significant modules for each contrast are shown. Size of circle indicates absolute effect size (AUC). Color of circle indicates significance (FDR < 0.05) and direction of effect (red = expression positively correlated with TRI, blue = negatively correlated). Absence of circle indicates non-significance.

2.4 Discussion

A meta-analysis of array and RNA-seq data revealed extensive transcriptomic response to Pandemrix vaccination in the **HIRD** cohort. At day 1, there was upregulation of genes and modules related to monocytes, interferon signalling, and the inflammatory response; and downregulation of T cell and **NK** cell genes and gene modules. Concordant changes in these gene modules were also reported by Nakaya *et al.* [158] at day 1 after MF59-adjuvanted seasonal **TIV** in young children, but changes in these modules were not as consistent in children who received non-adjuvanted **TIV**. The AS03 adjuvant in Pandemrix is thought to act by promoting chemokine secretion, predominantly targeting monocytes and macrophages [163, 248], which concurs with the strong upregulation of monocyte and **DC** modules observed at day 1 after Pandemrix. A large component of the expression response at day 1 may reflect response to the adjuvant. Most genes differentially expressed at day 1 returned to baseline expression by day 7. Nakaya *et al.* [158] saw a similar trend comparing day 0 and day 3 for MF59-adjuvanted **TIV**. Unadjuvanted seasonal **TIV** also causes peak transcriptomic induction at day 1 [153]. Although the timepoint resolution here is coarse, the early innate response to Pandemrix is transient, peaking less than 7 days, and likely less than 3 days post-vaccination. Upregulation of cell cycle, proliferating CD4⁺ T cell, and B (plasma) cell genes and modules were detected at day 7. This indicates a shift to the adaptive immune response, likely involving CD4⁺ T cell-supported differentiation and proliferation of **ASCs**.

Both day 1 and day 7 expression module changes were concordant with changes in cell populations seen in the **HIRD FACS** data. The greater magnitude of expression fold change of individual genes compared to cell abundance fold changes suggests the influence of both mechanisms [162]. Statistical adjustment for measured or estimated cell composition is one possibility; I explore these methods in **Chapter 3** and **Chapter 4**. An experimental approach would be *in vitro* stimulation of PBMCs with vaccine, ruling out cell migration, but not shifts in cell subtype composition [249].

The overall patterns of expression over time were consistent between array and RNA-seq, with the meta-analysis identifying genes with outlying effects in both platforms. In contrast, I was not able to replicate the 58 gene-level associations reported by Sobolev *et al.* [162] between day 7 expression and antibody response that were assessable in my meta-analysis. The difference was not wholly due to response definitions, as within the array data alone, switching from binary response status to **TRI** still replicated the majority of reported associations, but using either binary response status or **TRI** in the RNA-seq data alone found no significant associations. Initially, 15/58 signals replicated using frequentist random-effects meta-analysis to combine per-platform estimates. I do not consider these hits as robust, as the estimated between-platform heterogeneity was zero for all 15 of these signals. None of these signals replicated in the Bayesian random-effects meta-analysis, where prior information about τ could be incorporated, discouraging unrealistic estimates of zero heterogeneity. The Bayesian meta-analysis was in general more conservative, calling fewer differentially expressed genes compared to the frequentist analysis for all contrasts. Most of the 58 genes also had larger effects in the array dataset than in the RNA-seq dataset, possibly because the array data contains more extreme **TRIs**. At a single-gene level, significant associations with timepoint are robustly detectable, but associations with **TRI**

have effects too modest relative to the noise introduced by platform-dependent technical effects and dataset-dependent phenotype distributions.

Expression associations with antibody response were, however, observed at the gene set level, at modules associated with **TRI** as a whole. The strongest effects were observed at day 7, where modules related to adaptive immunity (cell cycle, stimulated CD4⁺ cells, plasma cells) were positively associated with **TRI**. These same modules were upregulated at day 7 compared to baseline; it seems that those individuals with the greatest antibody response to vaccination are most able to induce these modules by day 7 post-vaccination.

Module associations with **TRI** were also observed pre-vaccination with both positive (e.g. chemokines, proinflammatory DCs, B cells, platelet activation) and negative (e.g. monocytes) directions of effect, suggesting baseline immune state has influence on long-term antibody response to Pandemrix. Some of the positive associations have been previously reported for unadjuvanted seasonal influenza vaccines in multiple independent cohorts. The same B cell modules were reported by Nakaya *et al.* [157], and similar DC, inflammatory, and platelet activation modules were found to be predictive of antibody response in young adults [159]. The negative association of monocyte modules with antibody response at baseline was also reported by Nakaya *et al.* [157]. Interestingly, I detected the same negative associations at day 1 and day 7. Monocyte modules were one of the most upregulated modules at day 1, and although the module annotations do not separate monocyte subsets, abundance of CD16⁺ inflammatory monocytes was particularly increased at day 1 in the FACS data [162]. This lends some support to the hypothesis that chronic baseline inflammation or excessive/prolonged post-vaccination inflammation—specifically driven by monocytes—can be detrimental to the humoral response [157, 250, 251].

There are several caveats to consider when drawing comparisons to the systems vaccinology literature. Most studies are of unadjuvanted multivalent seasonal vaccines; **HIRD** used an adjuvanted monovalent pandemic vaccine. Most studies measure post-vaccination antibody response around the expected peak of day 28; **HIRD** measured later at day 63, which may attenuate the signal. The specific genes within modules driving associations may also differ between studies. Nevertheless, the ability to observe module-level associations with **TRI** also reported in previous studies with diverse populations, measurement platforms, influenza seasons, and analysis pipelines, is a stark contrast to difficulty of replicating single-gene associations even within the **HIRD** cohort itself. When the effect of individual genes on phenotype is expected to be subtle, module-level analyses are not only more sensitive, but appear to be more generalisable.

The next step is to explore the utility of the identified associations for prediction. Although I have identified highly significant associations between expression modules and antibody response, that does not imply the ability to accurately predict response from expression [161]—that is, the existence of molecular signatures. Some exploration can be done within **HIRD** using cross-validation, or by setting aside a subset (e.g. the array data) as a test set, but having an independent test set is especially important for prediction to guard against overfitting. Matched expression and antibody data are rare for adjuvanted and pandemic vaccines, so an initial effort would likely draw on published seasonal vaccine datasets (e.g. [157, 159]), with the aim of identifying shared molecular signatures.

The fundamental question of why gene expression and antibody responses vary between

HIRD individuals also remains. Which genes, if their expression were to be modulated, would lead to a change in antibody response? This is a critical question in the move from identifying correlates of protection and molecular signatures, towards targeted interventions to improve vaccine outcomes [146]. The descriptive design of the HIRD study does not lend itself to exploring causation between expression and antibody titres without a causal anchor. Interindividual genetic variation could play such a role; Chapter 3 will examine the impact of common host genetic variants on expression response in the HIRD cohort.

Chapter 3

Genetic architecture of transcriptomic response to Pandemrix vaccine

The work presented in this chapter is a collaboration between the Wellcome Sanger Institute, King's College London, the Francis Crick Institute, and the Biomedical Research Centre at Guy's and St Thomas' Hospital and King's College London. I would like to reiterate my thanks to the people and organisations mentioned at the beginning of Chapter 2.

3.1 Introduction

3.1.1 Host genetic factors affecting influenza vaccine response

Many human traits are heritable and complex—response to vaccination is no exception. Twin studies have demonstrated approximately 30–90 % heritability of antibody responses to many vaccines, including smallpox, hepatitis A and B, anthrax, pneumococcal, *Haemophilus influenzae* type b (Hib), diphtheria-tetanus-pertussis (DTP), and bacillus Calmette–Guérin (BCG) [252–255]. Candidate gene studies and genome-wide association studies (GWASs) have identified multiple genetic associations with antibody response [252, 253, 256, 257], including replicated associations for hepatitis B vaccine in a haplotype block in the human leukocyte antigen (HLA) region encompassing *HLA-DR* and *BTNL2*, and for measles vaccine in an intron of a receptor known to interact with measles virus, *CD46*.

In contrast, Brodin *et al.* [255] found anti-haemagglutinin (HA) antibody responses to seasonal influenza vaccine in 105 adult twin pairs (median age 44 yr) had no detectable heritability, alongside a general decrease in heritability of most immune parameters with age. They posited that the genetic contribution to response was overshadowed by environmental factors such as previous influenza vaccination or infection in adults, whereas the estimated heritability of the aforementioned vaccines was substantial because they are vaccines against non-circulating pathogens, or are childhood vaccines for which heritability was assessed in young children with shorter immune histories.

Nevertheless, a small number of candidate gene studies have identified genetic variants

associated with antibody response to influenza vaccines [257]. Gelder *et al.* [258] ($n = 73$) identified associations between HLA alleles in *HLA-DRB1* and *HLA-DQB1* with haemagglutination inhibition (HAI) seroconversion after trivalent inactivated influenza vaccine (TIV); Moss *et al.* [259] ($n = 185$) also found associations between HLA class II alleles (*HLA-DRB1*04:01* and *HLA-DPB1*04:01*) and HAI seroconversion after seasonal influenza vaccination. Poland *et al.* [260] ($n = 184$) tested HLA alleles, and single nucleotide polymorphisms (SNPs) in coding and regulatory regions of cytokine or cytokine receptor genes, for association with post-TIV HAI titres specific to H1 and H3 subtypes (two of the components of the trivalent vaccine). They reported nominally significant associations for two *HLA-A* alleles with H1-specific titres, six SNP associations with H1-specific titres and ten SNP associations with H3-specific titres. Egli *et al.* [261] ($n = 196$) identified a SNP upstream of *IFNL3* (rs8099917) to be associated with seroconversion post-TIV, and also found the SNP to be an expression quantitative trait locus (eQTL) for *IFNL3* expression in H1N1-stimulated peripheral blood mononuclear cells (PBMCs) in a second cohort ($n = 49$). Lastly, Avnir *et al.* [262] focused on a coding variant (rs55891010) in the part of *IGHV1-69* that encodes the complementarity-determining region (CDR) of broadly neutralising antibodies that bind influenza HA. One month after H5N1 avian influenza vaccination ($n = 85$), associations were detected with usage of *IGHV1-69* in the antibody repertoire, and serum antibody binding efficiency to H5N1 HA. The associations listed above have all been found in small cohorts and have not been validated by subsequent studies, so it remains unknown whether robust genetic associations with antibody response to influenza vaccines exist.

3.1.2 reQTLs induced by influenza vaccination

Host genetic variation could play a causal role in influenza vaccine response by altering the expression of genes as eQTLs. As described in Section 1.2.3 and Section 1.2.4, the effect sizes of eQTLs can be highly context-dependent, and many eQTLs in the immune system are response expression quantitative trait loci (reQTLs) only detectable after stimulation, not at baseline. One can map reQTLs considering vaccination as an *in vivo* immune stimulation. This usually involves measuring the transcriptome of immune cells before and after vaccination in genotyped individuals, then testing for genotype-dependent changes in expression. As expression is a key molecular intermediate between genotype and phenotype, a genotype-dependent change in expression after vaccination may be a mechanism mediating genotype-dependent antibody responses.

As reviewed in Section 1.2.4, few *in vivo* reQTL studies have been conducted, and even fewer studies have been conducted where the *in vivo* stimulation is vaccination, despite the potential for learning about genetic regulation of vaccine-induced expression responses. To my knowledge, there is only one such study: by Franco *et al.* [94] on response to seasonal inactivated TIV. Franco *et al.* [94] enrolled healthy Europeans adults into discovery ($n = 119$ males) and validation ($n = 128$ females) cohorts in two consecutive influenza seasons*. In each cohort, peripheral blood gene expression was measured by expression array on day 0 (baseline); and on days 1, 3, and 14 post-vaccination. Serum HAI and microneutralisation (MN) titres were measured against each of the three vaccine components at days 0, 14, and 28. The titre response index (TRI) [153] was computed from these titres as a single measure of antibody response adjusted for baseline titres.

*Sex-dependence of effects was not addressed.

Individuals were genotyped by genotyping array.

Cis-eQTL were mapped using a linear mixed model jointly over all four days, with day, genotype, day-genotype interaction, and a random intercept for individual as predictors; and gene expression the response variable. At 467 (non-independent) *eQTL* for 78 genes replicated in both cohorts, there was both a significant day effect (indicating the gene was differentially expressed post-vaccination) and a significant genotype effect (indicating the *eQTL* effect). To call *reQTLs*, *eQTLs* were also mapped separately for each day with a linear model including only genotype as a predictor, from which the model R^2 was computed as a rough measure of the variance in expression explained by the *eQTL* at each day. Franco *et al.* [94] then computed delta- R^2 : the maximum absolute deviation of the three post-vaccination R^2 s from the day 0 R^2 . Out of the *eQTLs* that replicated in both cohorts, 146 *eQTLs* for 34 genes ranking above the 99th percentile of the delta- R^2 distribution were defined as *reQTLs*. The union of the 78 and 34 genes from the above analyses (98 genes with differential gene expression (*DGE*) and an *eQTL*; or a *reQTL*) was enriched for pathways and gene sets related to antigen processing and presentation, CD8⁺ T cell-mediated apoptosis, dendritic cell (DC) maturation and function, and membrane trafficking. Lastly, integrating antibody titre data, they filtered down to 20 genes with expression correlated to *TRI* at any day, with an *eQTL*, and with either post-vaccination differential expression *or* a *reQTL* effect. Seven genes out of these 20 were involved in antigen transport, processing, or presentation in antigen-presenting cells (APCs): *NAPSA*, *C1orf85*, *GM2A*, *SNX29*, *FGD2*, *TAP2*, and *DYNLT1*.

Critically, Franco *et al.* [94] recognised that just assessing overlap of multiple filtering criteria cannot infer the direction of causal relationships between genetic variation, expression and *TRI*. They attempted a model comparison with the CIT [263] method to resolve the directionality of association between expression and *TRI*, finding suggestive evidence of causal effect on *TRI* mediated by expression at several *eQTL*. Unfortunately, they also evaluated that the power of the CIT was only ~60% at their total sample size of $n = 247$. Nevertheless, the study is proof of concept that integration of genotype, expression, and antibody response data in an *in vivo* *reQTL* framework can identify genes under genetic regulation likely to be involved in vaccine response.

3.1.3 Chapter summary

The Human Immune Response Dynamics (HIRD) cohort represents a unique opportunity for detecting host genetic contributions to influenza vaccine response. Similar to Franco *et al.* [94], expression, antibody response, and genotypes are all available for the same individuals. As Pandemrix is against a pandemic strain that had not been in seasonal circulation for decades at the time of cohort recruitment, responses will be less driven by individual immune history, so power to detect genetic associations is expected to be greater. In Chapter 2, I characterised differential expression induced by Pandemrix, as well as expression associations with antibody titres. In this chapter—given that HIRD is too small for a direct *GWAS* of antibody response—I focus on the genetic contribution to expression response. I apply the *in vivo* *reQTL* framework, aiming to characterise the association of common genetic variants with expression across multiple timepoints, and pinpoint genes important to Pandemrix response.

3.2 Methods

3.2.1 Overall strategy for reQTL mapping

A plethora of approaches to mapping eQTLs with linear models exist; each approach has its own advantages, disadvantages, and assumptions. When the task is also to define reQTL between multiple conditions, the diversity of possible approaches further multiplies. Here I will discuss aspects of the data and available methodologies that led to the final modelling strategy adopted in this chapter.

3.2.1.1 Adjusting for population structure using linear mixed models

Population structure occurs when the samples in a study are not independent, but structured due to genetic relatedness. Genetic association studies assume that the individuals in a sample are unrelated (or at least sufficiently distantly related) [264–266]. Relatedness, and thus population structure, occurs at different scales. Population stratification refers to systematic differences in allele frequencies and genetic background between human populations due to demographic history. This represents large-scale structure where individuals are related due to shared ancestry [188, 265]. At a smaller scale, sample individuals can be related due to being in the same family. The presence of more relatedness in a sample than is assumed is the problem of cryptic relatedness [264–266]; this can be at any scale, but more often the term refers to recent relatedness.

In the context of eQTL mapping (and genetic association studies in general), where the aim is to assess the effect of a single genetic variant on expression, there is potential for confounding. The issue (well-reviewed in [266, 267]) is that we fit a marginal model to estimate the effect of a single variant x_k on the phenotype y :

$$y = \mu + \beta_k x_k + \epsilon \quad (3.1)$$

where μ is the intercept, and $\epsilon \sim N(0, \sigma_e^2 I)$ is the error term that represents environmental and stochastic sources of variation. The variance-covariance matrix for error term is a scalar matrix, encoding the classic regression assumptions of homoscedasticity and uncorrelated errors. A more appropriate data generating model is:

$$y = \mu + \beta_k x_k + G + \epsilon \quad (3.2)$$

where $G = \sum_{i \neq k} \beta_i x_i$ represents the effect of the genetic background at all other variants in the genome. As many variants can be expected to affect a complex polygenic trait, G has some causal effect on y . Population structure means there can be a shared cause of G and x_k such as ancestry. This opens a backdoor path $x_k \leftarrow \text{ancestry} \rightarrow G \rightarrow y$, confounding the relationship between x_k and y . In Eq. (3.1), when one estimates the coefficients, the effects of the omitted variable G will be attributed to x_k , resulting in spurious associations and genomic inflation of test statistics [188]. Here, G represents exactly the confounding due to genetic background, but there are other possible confounders, such as shared environmental factors that differ systematically between populations [268]. A popular approach to avoid confounding is to include genotype principal components (PCs) as fixed effects in the regression [185, 269], thus blocking the backdoor path

from x_k to y . Genotype PCs represent population stratification effects like ancestry, but also act to block confounding from genetic background and environmental effects by proxy [268].

Unfortunately, genotype PCs alone cannot account for smaller-scale population structure [188]. An approach that can explicitly model such population structure is the linear mixed model (LMM) [188, 267, 269], which expresses the idea that more genetically correlated individuals are expected to be more phenotypically correlated [268]. A typical model form is:

$$y = \mu + \beta_k x_k + u + \epsilon \quad (3.3)$$

where random effect $u \sim N(0, \sigma_g^2 K)$ has a variance-covariance matrix proportional to the genetic correlation between individuals: the kinship matrix K . This improves on Eq. (3.2) by recognising that the variants in G are correlated. σ_g^2 is often called the (genetic) variance component; the larger it is, the more phenotypic variance is explained by genetic background [267]. Although LMMs were originally developed in the context of animal breeding, where K is computed from a known pedigree, it can also be computed from genome-wide SNP data [266, 269]. Unlike pedigree-based kinships that range from zero (unrelated) to one (self or identical twin), SNP-based relatedness values represent average correlations of alleles between individuals [270], hence may be negative or greater than one [271]. This does not affect their usage in LMMs.

The HIRD genotype data has already been filtered such that no pair of individuals are first-degree relatives or closer (Section 2.2.4), but cryptic relatedness may remain. The multi-ethnicity of the cohort means there is large-scale population structure from ancestry (Fig. 2.5). Genotype PCs were computed to represent axes of variation due to ancestry. These were included as covariates in DGE analyses in Chapter 2 to improve efficiency by explaining some variation in expression. For eQTL mapping in this chapter, I use both a random effect in an LMM and PC fixed effects to correct for population structure. This may not be strictly necessary if the random effect can correct for large-scale structure, but does not seem to impact power or type I error rate [272], and may have some benefit at SNPs with very different allele distributions between populations (unusually differentiated) [188].

The performance of various software implementations for kinship estimation from genome-wide SNP data and LMMs are highly comparable; the specific choice of implementation can usually be made on the basis of computational efficiency [269]. In this chapter, I use LMMs implemented in LIMIX [273] with kinship matrices estimated by LDAK [274].

3.2.1.2 Multi-condition models

Since the aim of this chapter is to identify genetic variation that affects expression response to vaccination, it may seem most direct to model the change in each individual's expression after vaccination as the response variable. This approach has been applied for identification of condition-specific eQTL, typically with the response taking units of log fold change between conditions (e.g. [275–277]). Although potentially powerful if eQTL effects are small and opposite between conditions [276], it is analogous to the “change score” approach, which can suffer from regression to the mean, and increased uncertainty from the variance sum law if effects between conditions have positive covariance [171, 278]. Instead, I map eQTLs within each of three

timepoint conditions (day -7/0 pre-vaccination baseline, day 1, and day 7), and find reQTLs by looking for eQTLs that have different effects between conditions. As in Section 2.2.9.2, day -7 and day 0 array measurements were treated as repeated measures of the baseline timepoint. Unlike a test for difference implemented using a genotype-condition interaction term in a joint regression model, homoscedasticity of errors is not assumed for all conditions [279].

Within each timepoint, the HIRD dataset includes expression measured by both array and RNA sequencing (RNA-seq). As discussed in Section 2.2.9.3, it is difficult to directly estimate the between-study heterogeneity when the number of studies is small, thus Bayesian meta-analysis was preferred for combining array and RNA-seq DGE estimates. That method does not scale to eQTL analysis, where the number of tests is large, in the order of thousands of tests per gene, versus the handful DGE contrasts per gene performed in Chapter 2. Instead, I perform a mega-analysis within each timepoint, first merging array and RNA-seq expression estimates into a single matrix with ComBat [211]. For comparison purposes, eQTL analyses were also run in the array and RNA-seq samples separately.

Defining whether an eQTL is shared between conditions can be a tricky business. Naively, after mapping eQTLs separately in each condition, one can assess the overlap of significant associations between conditions. This underestimates sharing due to the difficulty of distinguishing true lack of sharing from missed discoveries as a consequence of incomplete power within each condition [68, 280]. Condition-by-condition analysis also cannot borrow information across conditions for mapping shared associations [280–282]. Counterintuitively, a joint multivariate analysis may be more powerful even when associations are not shared across all conditions [283].

A variety of models have been employed for joint eQTL mapping, including the use of classical multivariate methods such as multivariate analysis of variance (MANOVA) [80], frequentist meta-analyses (e.g. Meta-Tissue [284], METASOFT [285]), and Bayesian models (e.g. eQtlBma [280], MT-HESS [286], MT-eQTL [287]). Joint mapping has repeatedly been demonstrated to be more powerful than condition-by-condition analysis, and recent joint methods are now computationally efficient when scaling to large numbers of conditions and variants tested (e.g. RECOV [288], mashr [281], HT-eQTL [282]). In this chapter, I apply mashr [281] for the joint estimation of eQTL effects across my three timepoints. The method learns patterns of correlation among multiple conditions empirically from condition-by-condition summary statistics, then applies shrinkage to provide improved posterior effect size estimates, and computes measures of significance per condition.

3.2.1.3 Additional expression preprocessing

There are a number of transformations often applied to expression data before eQTL mapping, such as the rank-based inverse normal transformation (INT) (e.g. GTEx v8 [54]). INTs work by matching sample quantiles to quantiles of the standard normal distribution, which conforms often non-normal expression data to an approximately normal distribution, improving computation speed in large samples [289] and reducing the impact of expression outliers. In the context of genetic association studies, the practice of applying rank-based INT to phenotypes has been criticised for only guaranteeing approximate normality of residuals when effect sizes are small, and potentially inflating the type I error in linear models that include interaction terms [290].

More recent simulations suggest that **INTs** have good power and type I error control across commonly-encountered distributions of non-normal residuals [289]. Another common transform is standardising (centering and scaling to zero mean and unit variance e.g. eQTLGen Consortium [58]), often done so that effects across genes and studies can be comparably interpreted in units of standard deviation expression [291]. In multi-condition datasets, data transformations are typically applied within conditions (e.g. within each tissue individually in GTEx v8 [54]).

Simulations were performed to evaluate the effect of the aforementioned transformations on **reQTL** detection between a hypothetical day 0 baseline and day 1 post-vaccination condition. The size of a **reQTL** effect depends on the scale of the expression data; here I define the size of a **reQTL** as the difference in **eQTL** slopes (betas) for the same variant-gene pair between conditions with expression measured on the \log_2 scale. The boxed facets in Fig. 3.1 represent undesirable effects of transformations on **reQTL** effect sizes. Rank-based **INT** induces false shared **eQTL** effects between conditions in scenarios 4 and 5 (e.g. row “Rank-based INT” in Fig. 3.1). In general, transformations that scale within condition are not appropriate, as different variances within conditions contribute to the **reQTL** effect (e.g. “Scale within day”). Scaling without separating conditions is also problematic, since the total variance also affects the **reQTL** effect size. For example, in “Scale”, scenarios 2 and 4 have the same 1 unit increase in beta pre-transformation (the same fold-change between conditions), but after scaling-only the beta increases are $0.75 - 0 = 0.75$ and $0.8 - 0.4 = 0.4$ respectively—scenario 4 now looks like a **reQTL** of weaker effect.

In light of these simulations, I decided that neither rank-based **INT** nor standardisation were appropriate. Only the centering-only transformations (e.g. “Center”) avoided both false shared effects and preserved relative **reQTL** effect sizes. The simple inclusion of an intercept term in the **eQTL** model already achieves this, so no additional expression transformations were applied before **eQTL** mapping. Not performing a rank-based transform does lose the advantage of reining in outliers, but the expression data have already been preprocessed to remove low-expression outliers in Section 2.2.7. Many other preprocessing steps done prior to this stage in the pipeline (e.g. variance-stabilisation, ComBat batch effect correction) are also expression transformations, but I only consider the preservation of **reQTL** effects defined from expression values post-adjustment for technical effects to be important, so I did not consider those steps in my simulations.

3.2.2 Genotype phasing and imputation

Genotyping and pre-imputation processing are described in Section 2.2.3 and Section 2.2.4. Prior to imputation, 213 277 monomorphic variants that provide no information for imputation were removed. Variant alleles were aligned such that the reference allele matched the GRCh37 reference, and 358 indels were removed, leaving only **SNPs**. Imputation for the autosomes and X chromosome was conducted using the Sanger Imputation Service*, which involved pre-phasing (separate estimation of haplotypes before imputation to improve imputation speed) with EAGLE2 (v2.4) [292] and imputation with PBWT (v3.1) [293] against the Haplotype Reference Consortium (r1.1) panel [294]. Imputed **SNPs** were lifted-over from GRCh37 to GRCh38 coordinates using CrossMap [295]. Poorly-imputed **SNPs** with imputation information score $\text{INFO} < 0.4$ were

*<https://www.sanger.ac.uk/tool/sanger-imputation-service/>

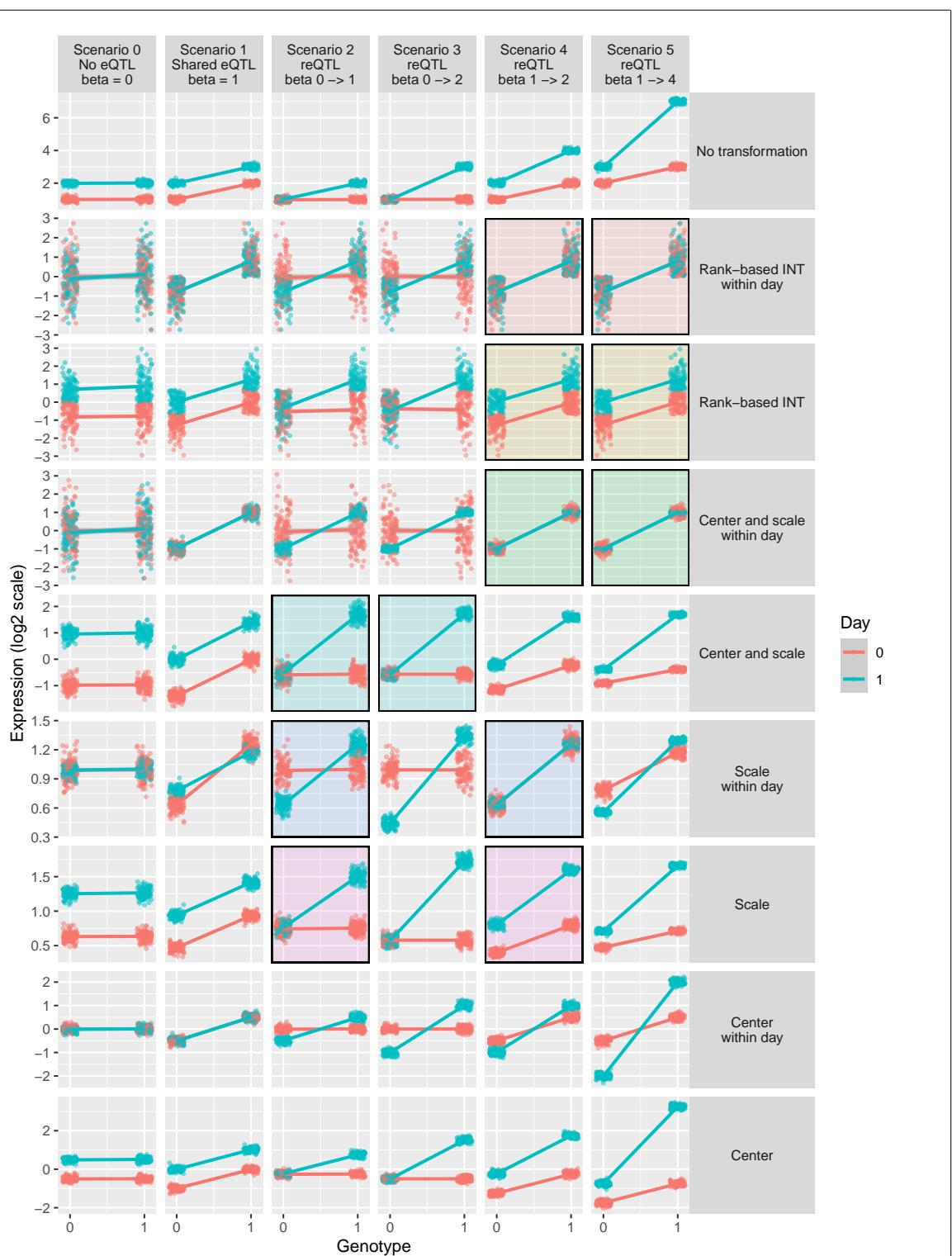


Figure 3.1: Simulating the effect of data transformation on reQTL effects. Expression values on the log scale were simulated for 200 individuals (100 with each genotype) at a day 0 baseline and day 1 post-vaccination timepoint. Gene expression is upregulated at day 1 by $\log_2 \text{FC} = 1$. Six scenarios were simulated with different gene-variant pairs (columns) corresponding to different eQTL and/or reQTL effects between day 1 and day 0; the size of the reQTL effect (difference in beta between day 1 and day 0) was set to 0, 0, 1, 2, 1, and 3 for the six scenarios. Gaussian noise with mean = 0 and standard deviation = 0.1 was added. The top row is the ground truth. In following rows, a different transformation was applied within each row: a rank-based INT (Blom offset for fractional ranks [289, 290]), standardising (centering and scaling), centering only, or scaling only. Highlighted pairs of scenario-transform combinations on each row represent false positives or negatives where the size of the relative reQTL effects are no longer correct.

removed, leaving 40 290 981 SNPs measured for the 169 genotyped individuals.

3.2.3 Estimation of kinship matrices

When testing a variant for association using LMMs, the kinship matrix used should not include that variant to avoid loss of power from “proximal contamination” [296]. A simple way to avoid this is to compute a leave-one-chromosome-out (LOCO) kinship matrix using all variants except the ones on the tested variant’s chromosome [297]. I estimated kinship in the HIRD data from common autosomal variants, using LDAK (v5.0) [274], which computes SNP-based kinship matrices weighting SNPs by linkage disequilibrium (LD) and accounting for genotype uncertainty. Filtered pre-imputation sample genotypes from Section 3.2.2 were pruned to **MAF** > 0.05. A kinship matrix was computed for each autosome, then combined into a single genome-wide matrix using LDAK `--join-kins`. To obtain a LOCO kinship matrix for each autosome, each autosome’s kinship matrix was then subtracted from this genome-wide matrix (LDAK `--sub-grm`).

3.2.4 Estimation of cell type abundance from expression

PBMC samples are a mixture of immune cells, and a fixed input of RNA extracted from that mixture is used to estimate expression, so estimates for genes that have cell type-specific expression depend on the relative abundances of each cell type in each sample. Sobolev *et al.* [162] showed these abundances shift after Pandemrix vaccination. As genotype can be assumed to stay constant, it is valid to compare the effect size of genotype on expression between multiple timepoints to call reQTLs, but changes in cell type abundance complicate this by modifying both expression (i.e. cell type-specific expression) and the effect of genotype on expression (i.e. cell type-specific eQTL effects). Immune cell abundance also varies naturally between healthy individuals [109, 255], so it is important to model these effects not only post-vaccination, but also at baseline.

Cell type abundance directly measured via fluorescence-activated cell sorting (FACS) was only available for a small subset of HIRD individuals (Section 2.2.1), so I computed cell type abundance estimates from the expression data as an alternative. *In silico* estimates have previously been used as covariates for eQTL analyses in bulk samples where cell type-specific effects are expected [71, 72, 74, 96]. As the estimates are based on the expression of multiple genes, it is not entirely circular to use them as covariates in this way for per-gene eQTL models. I selected `xCell` [298], which has been shown to outperform other deconvolution methods for cell type-specific eQTL mapping in blood [74]. `xCell` computes enrichment scores based on the expression ranks of approximately 10 000 signature genes derived from purified cell types, works for both array and RNA-seq expression data, and implements “spillover compensation” that reduces dependency of estimates between related cell types [298]. `xCell` was originally developed for tumour samples, so many of the built-in cell types are not expected to be in PBMCs. Reviewing the literature to find which broad classes of peripheral blood cell types are commonly expected in the PBMC compartment [96, 299, 300], I selected 7/64 of the built-in cell types: CD4⁺ T cells, CD8⁺ T cells, B cells, plasma cells, natural killer (NK) cells, monocytes, and DCs. Array and RNA-seq data from Section 2.2.8 and Section 2.2.7 were processed separately, as different internal `xCell` parameters are used for each platform. The large batch effect present in the array expression was

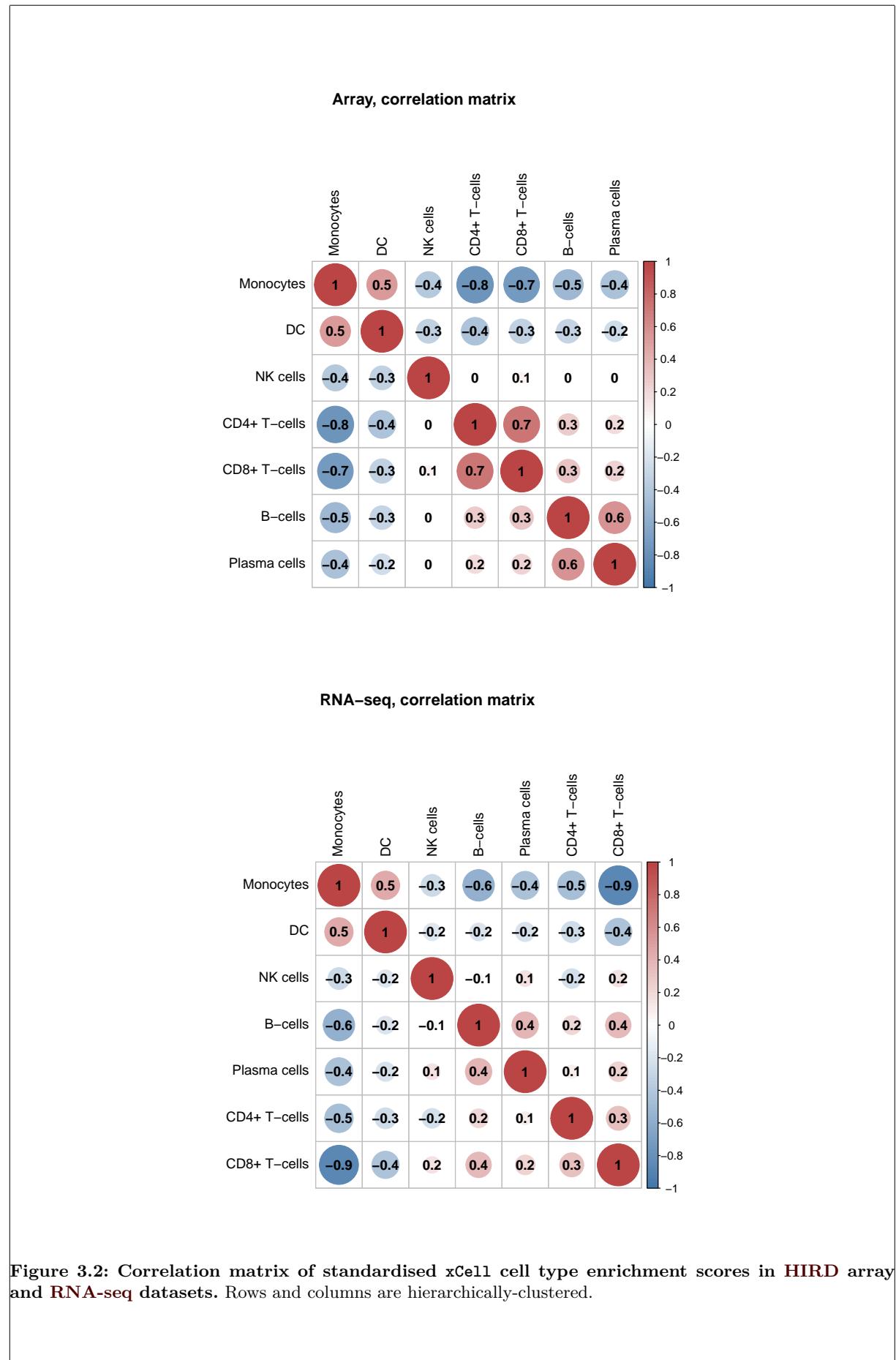
first removed using ComBat [211]. Finally, enrichment scores were standardised across timepoints, so that a score of zero represents the average abundance of that cell type across all timepoints.

As with actual cell type abundances, **xCell** enrichment scores are correlated (Fig. 3.2). Imprecise coefficient estimates due to multicollinearity may be a problem when these scores are included as independent variables in **eQTL** models*. To select a subset of cell type scores, I performed a principal component analysis (PCA) of the cell type scores separately in array and **RNA-seq** datasets (to prevent axes reflecting platform rather than cell type), then determined the number of **PCs** that exceeded the eigenvalues-greater-than-one rule of thumb [302]. In both array and **RNA-seq** datasets, this number of **PCs** was three. The cumulative percentage of variance explained by the top three **PCs** was 81.02% and 74.58% in the array and **RNA-seq** datasets respectively. Since the **PCs** between the array and **RNA-seq** are not directly comparable, I selected three cell types with high contributions to the top three **PCs** in both datasets: monocytes, **NK** cells, and plasma cells (Fig. 3.3)—Sobolev *et al.* [162] also reported monocytes and plasma cells to be the cell types with the highest abundance increases at days 1 and 7 respectively. Additionally, using the actual cell type scores rather than **PCs** as covariates provides more interpretable regression coefficients for those terms.

Scores were validated against **FACS** measurements from Sobolev *et al.* [162] in the subset of ~40 individuals that had both expression and **FACS** data. Depending on each **FACS** panel's gating strategy for each cell subset, the data were in units of either absolute counts or percentage of the previously gated population. Values were normalised by rank-based **INT** within each panel and cell subset (Astle *et al.* [303] took a similar approach for cell abundance data using a quantile-based **INT**). Missing values were then imputed with MissForest [304], a random forest imputation method suitable for high-dimensional mixed-type data where $p \gg n$. The method establishes an initial guess for missing values using mean- or mode-imputation, then a random forest is trained on the observed part of the data and used to predict and update the values of the missing part. The process repeats iteratively until convergence.

Although the increases in **xCell** score for monocytes at day 1 and plasma cells at day 7 do reflect the increases in those cell types observed by Sobolev *et al.* [162], overall correlation between **xCell** and **FACS** was poor (Fig. 3.4). Substantial discrepancy is expected, as the cell types as defined in the **xCell** signatures do not directly correspond to the combinations of surface markers used for **FACS**; the comparison is against the closest match. The **FACS** gating strategy also meant that for some cell populations, the only available **FACS** measure was a proportion of the previously gated population, whereas **xCell** attempts to estimate scores that represent enrichments in the whole mixture. The accuracy of the built-in signatures may also be lower when applied to the expression matrix for a stimulated state, where an enrichment-based method may not be able to distinguish per-cell differential expression of signature genes from changes in cell abundance. A custom signature matrix can be used for **xCell**, perhaps drawn from an independent study with similar stimulation conditions as **HIRD**, such as Franco *et al.* [94], but this would not solve the issue of coupled differential expression and cell abundance. Weighing

*High correlation between predictors is not necessary nor sufficient by itself to induce multicollinearity (predictors being linearly-related), but multiple correlation (how well predictors can be predicted as linear combinations of other predictors) does have an inverse relationship with the standard error of coefficient estimates [301].



Array, contributions of each variable

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
CD4+ T-cells	0.19	0.01	0.23	0.01	0.01	0.31	0.25
CD8+ T-cells	0.19	0.01	0.17	0.07	0	0.56	0.01
B-cells	0.12	0.21	0.1	0	0.54	0	0.03
Plasma cells	0.09	0.3	0.12	0.01	0.45	0	0.03
NK cells	0.02	0.39	0.32	0.15	0	0	0.12
Monocytes	0.27	0.02	0	0.06	0	0.09	0.56
DC	0.13	0.06	0.06	0.7	0	0.04	0

RNA-seq, contributions of each variable

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
CD4+ T-cells	0.09	0.31	0.09	0.02	0.34	0.13	0.04
CD8+ T-cells	0.23	0	0.06	0.16	0.01	0.26	0.27
B-cells	0.14	0.03	0.29	0.05	0.17	0.29	0.03
Plasma cells	0.1	0.03	0.37	0.25	0.13	0.11	0.02
NK cells	0.02	0.62	0.02	0.02	0.1	0.19	0.03
Monocytes	0.31	0	0	0.06	0.01	0.01	0.61
DC	0.12	0.01	0.16	0.45	0.24	0.01	0

Figure 3.3: Contribution of each cell type score to each PC dimension after PCA of standardised xCell cell type enrichment scores. Contribution is calculated as the squared correlation between a variable and a PC (\cos^2), scaled to the sum of \cos^2 for all variables with that PC. High contributions indicate variables that are highly correlated with the PC.

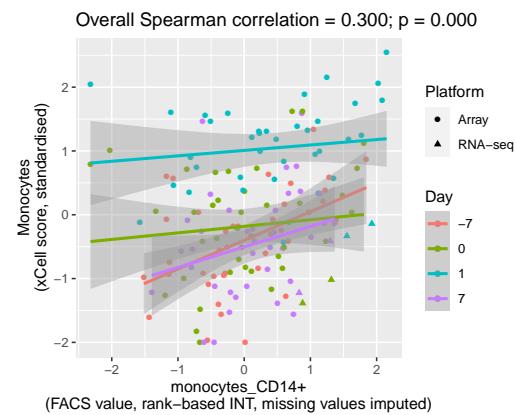
the downsides of having imperfect estimates of cell type abundance against the downsides of not accounting for abundance, or excluding samples without FACS measurements, I chose to continue the analysis using the `xCell` scores. These scores can distinguish large changes in cell abundances between days, but may not be reliable for distinguishing small differences in abundance between individuals within a timepoint.

3.2.5 Finding unmeasured covariates using factor analysis

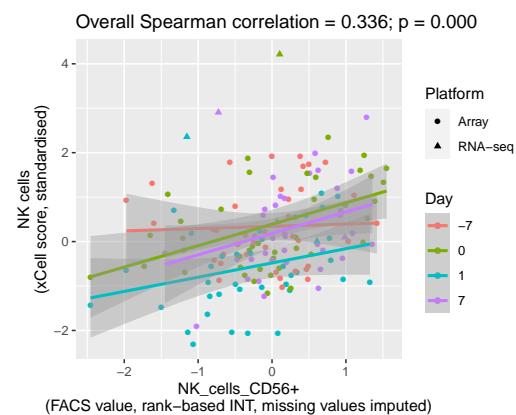
Apart from cell type abundance, a myriad of unmeasured variables contribute to expression variation. Hidden determinants of expression variation were learnt using PEER [182]. As suggested by Stegle *et al.* [182], I used `DESeq2::vst` [237] to perform between-sample normalisation and variance stabilisation on the RNA-seq count data*. ComBat [211] was then applied to merge array and RNA-seq data into a single log scale expression matrix per timepoint, treating the largest global effects on expression—the two array batches and three RNA-seq library preparation pools (Fig. 2.13)—as known batch effects. Given a set of known covariates (intercept, sex, four genotype PCs from Section 2.2.5 representing ancestry, and the three `xCell` scores estimated above in Section 3.2.4), PEER was used to estimate additional hidden factors that explain variation in the expression matrix. These can be technical (e.g. sample quality/concentration, library preparation plate/reagents, processing time, lane/flow cell) or biological (e.g. cell type composition, ancestry). Factors are assumed to be unmeasured covariates that have global effects on a large fraction of genes, whereas a *cis*-eQTL will typically only have local effects, so including factors as covariates should not introduce dependence with the genotype term, but should explain some of the residual variation, improving power to detect *cis*-eQTLs. The analysis was run per timepoint, otherwise global changes in expression between timepoints induced by the vaccine would be recapitulated as factors.

Correlating the estimated factors to a larger set of known covariates revealed many correlations with `xCell` estimates, indicating that cell type abundance does indeed have substantial global effects on the expression matrix (Fig. 3.5). These factors likely represent additional cell types with abundances that have a global effect on the expression matrix, and when used as covariates in combination with the three major cell type scores selected in Section 3.2.4, should improve overall adjustment for cell composition. There was little correlation with known array or RNA-seq batch effects, indicating ComBat did an adequate job of removing batch- and platform-dependent global effects on expression prior to PEER. Note that I did not leave this adjustment for PEER to perform, as ComBat estimates centering and scaling factors per gene and batch, whereas the use of PEER factors represents a mean-only per-gene adjustment. Given the severity of the batch effect in this dataset, especially between platforms, mean-only adjustment may be insufficient [214], particularly in the context of *cis*-eQTL mapping where associated variants will only explain a small fraction of expression variance.

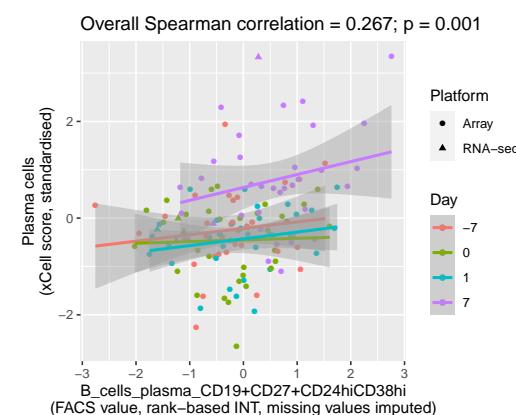
*The count data were taken from Section 2.2.7 before trimmed mean of M-values (TMM) normalisation and `limma::voom` transformation, as PEER cannot use the weights output by those methods for between-sample normalisation and variance stabilisation as `limma` can.



(a) Monocytes.

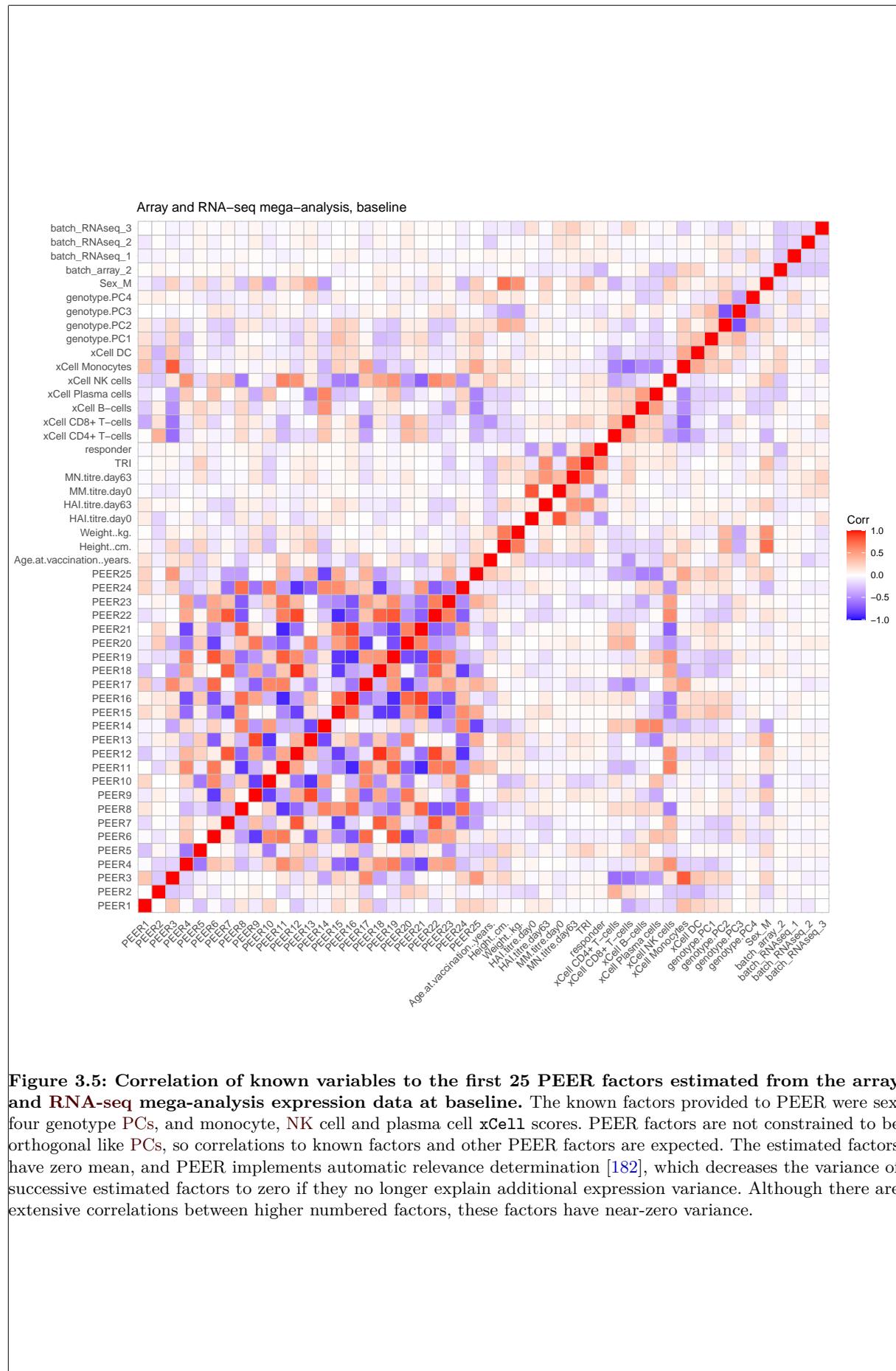


(b) NK cells.



(c) Plasma cells.

Figure 3.4: Comparison of standardised xCell scores with normalised HIRD FACS measurements, for monocytes, NK cells, and plasma cells. The comparisons are against the most comparable measurements available in the FACS data of Sobolev *et al.* [162]: CD14⁺ monocyte count, CD56⁺ NK cell count, and the proportion of CD19⁺ B cells that were CD19⁺CD27⁺CD24^{hi}CD38^{hi} plasma cells. Missing FACS values were first imputed with MissForest after rank-based INT transformation.



3.2.6 eQTL mapping per timepoint

I mapped **eQTLs** within each timepoint using LIMIX [273], which implements univariate and multivariate **LMMs** with one or more random effects. Imputed genotype probabilities were converted to continuous alternate allele dosages using bcftools (1.7-1-ge07034a)*. Variants with sample **AC** < 15 of the minor allele within each timepoint were excluded, corresponding to a 5–7% **MAF** depending on sample size (145 at baseline, 105 at day 1, and 107 at day 7). At these sample sizes, **false discovery rate (FDR)** cannot be controlled by standard hierarchical **FDR** methods without a **MAF** filter of approximately 5–10% [305].

At each of 13 570 genes, at all *cis*-variants within within ± 1 Mbp of the gene **transcription start site (TSS)**, I fit the following model to map **eQTLs**:

$$y = 1 + \text{sex} + \sum_{i=1}^4 \text{PC}_i + \sum_{i=1}^3 \text{xCell} + \sum_{i=1}^k \text{PEER}_i + \beta x + u + \epsilon \quad (3.4)$$

where the **eQTL** effect size is the slope of the genotype fixed effect β , the average additive effect of the alternate allele [13]; and $u \sim N(0, \sigma_g^2 K)$ is a random effect with zero mean and covariance matrix proportional to the **LOCO** kinship matrix for variant x . For chromosome X variants, no **LOCO** matrix was available from LDAK, so the matrix for chromosome 1 was used. Known covariates and PEER factors from Section 3.2.5 were included. PEER factors are automatically weighted such that the variance of factors tends to zero as more factors are estimated, hence continuing to add more and more factors as covariates will not continue to improve **eQTL** detection power, and eventually the model degrees of freedom will be depleted. To optimise the number of factors k to include[†], per-timepoint **eQTL** mapping was performed in chromosome 1, iteratively increasing the number of factors until including additional factors provides no further benefit, and the number of **eQTLs** detected stabilises. I settled on a final choice of $k = 10$ factors for baseline, 5 factors for day 1, and 5 factors for day 7 (Fig. 3.6).

3.2.7 Joint eQTL analysis across timepoints

Joint analysis was conducted with **mashr** [281] at 40 197 618 gene-variant pairs (mean of 2962 tests per gene) for which summary statistics from within timepoint mapping were available in all three timepoint conditions. For n conditions, the **mashr** model incorporates multiple $n \times n$ canonical and data-driven covariance matrices to represent patterns of effects across conditions. Canonical matrices include the identity matrix (representing independent effects between conditions), singleton matrices (effects only in one condition), a matrix of ones (equal effects in all conditions), and other patterns of correlations. Data-driven covariance matrices represent patterns of effect observed empirically, derived from dimension reduction of a strong subset of tests likely to have an effect in at least one condition. I took the most significant variants per gene per condition, which ensures strong condition-specific effects are included, then further filtered to only nominally significant tests, resulting in a strong subset of 45962 tests used

*<https://samtools.github.io/bcftools/>

[†]I avoid the commonly performed two-stage approach of using PEER residuals as expression phenotypes, as the degrees of freedom for the **eQTL** model will be incorrect, which can have a substantial effect on estimates at this modest sample size.

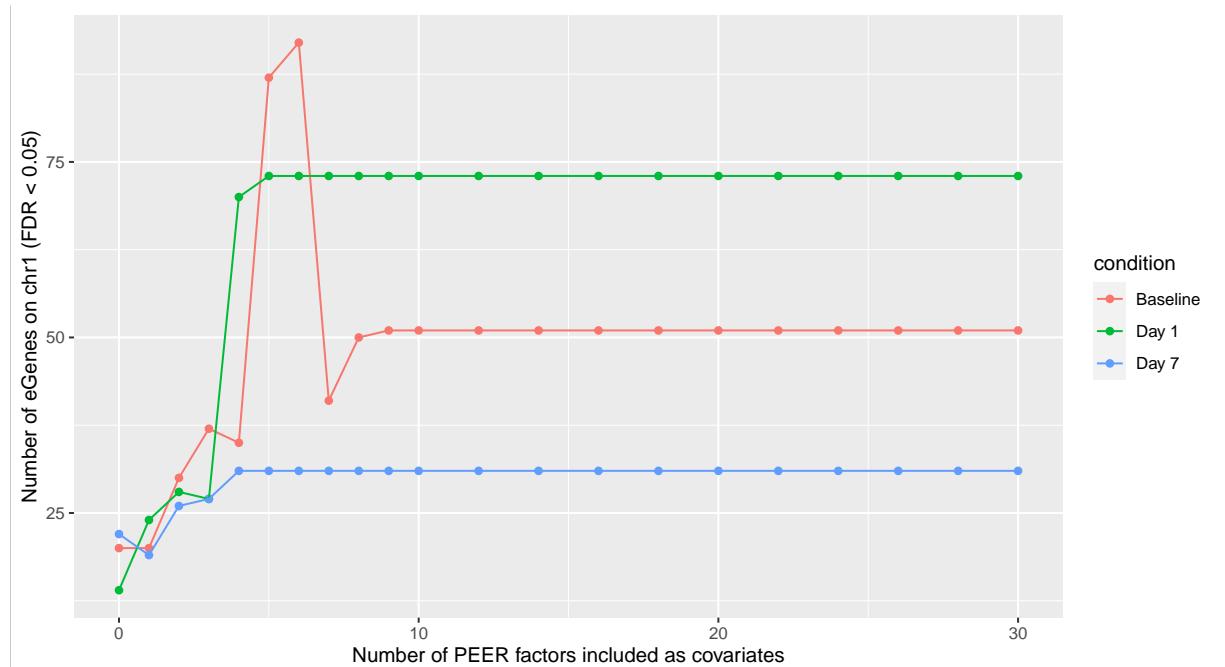


Figure 3.6: Number of significant genes with an eQTL detected on chromosome 1 as a function of the number of PEER factors included as covariates. FDR computed with hierarchical Bonferroni-Benjamini-Hochberg (BH) [305] with significance threshold set at 0.05. The number of eGenes stabilises since higher number PEER factors explain less and less variance in expression, thus having less and less influence on the regression.

to calculate data-driven covariance matrices.

The `mashr` model was trained on a random subset of 200 000 tests, using the exchangeable Z (EZ) model (assumes effects are independent of their standard errors, which performed better in GTEx data [281]). The correlation of null tests between conditions—critical to account for due to the repeated measures structure of the data—was estimated using `mashr::estimate_null_correlation`, which uses tests from the random subset that have small absolute z -scores. The fitted model was used as a prior to compute posterior effects and standard errors for all tests through shrinkage. A condition-specific Bayesian measure of significance is returned: local false sign rate (LFSR), which gives the probability that the declared sign of the effect is incorrect [238]. Note that `mashr` is the multiple-condition extension of `ashr` [238], previously used in Section 2.2.9.7 for computing posterior effects and their significance in DGE analyses.

3.2.8 Defining shared eQTLs and reQTLs

Many of the tested variants for each gene will be in high LD. To unambiguously select a lead eQTL variant per gene for comparison across timepoints, I selected the variant with the lowest LFSR over all conditions. If multiple variants had that same lowest LFSR value, ties were broken by highest imputation INFO, highest MAF, most upstream of the TSS, and finally genomic coordinate. Ties were not frequent. Sharing was then evaluated for that gene-variant pair across all three conditions.

Thresholding on the LFSR is not appropriate for determining sharing, as the difference

between significant and non-significant effect estimates in two conditions is not necessarily significant [306, 307]. Urbut *et al.* [281] provides a heuristic that two effects are shared by magnitude if they have the same sign, and are also within a factor of two, but this does not consider the posterior standard error of the estimates. Between a pair of effects in two conditions x and y , I computed a z -statistic for the difference in effects:

$$z = \frac{\beta_x - \beta_y}{\sqrt{\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}}} \quad (3.5)$$

This is a common strategy for comparing regression coefficients [279, 306] and has also been applied to call reQTLs by Kim-Hellmuth *et al.* [85]. Like Kim-Hellmuth *et al.* [85], I assume the posterior pairwise covariance of effects σ_{xy} is zero. This is conservative if the covariance is actually positive. A Wald test p -value for the difference can be computed, as under the null hypothesis of zero difference, asymptotically $z \sim \mathcal{N}(0, 1)$. I use nominal $p < 0.05$ as a heuristic threshold to separate shared and reQTL effects, and also computed the corresponding BH FDR per timepoint as a formal measure of significance. Note that even a nominal $p < 0.05$ threshold is still more stringent than calling sharing using the LFSR = 0.05 as a threshold (e.g. [74, 91]) or the 2-fold difference in magnitude threshold suggested by Urbut *et al.* [281].

Another statistic that quantifies the strength of an eQTL is the proportion of variance explained (PVE) by the variant. This was approximated using the following formula from Shim *et al.* [308] for variant X and expression Y :

$$\text{PVE} = \frac{\beta^2 \text{Var}(X)}{\text{Var}(Y)} = \frac{\beta^2 \text{Var}(X)}{\beta^2 \text{Var}(X) + \sigma_e^2} \approx \frac{\beta_p^2 2pq}{\beta_p^2 2pq + \sigma_p^2 2Npq} \quad (3.6)$$

where β is the beta from a simple linear regression of Y on X , σ_e^2 is the residual error, β_p is the posterior beta from mashr, σ_p is its posterior standard error from mashr, N is the sample size, p is the sample MAF, and $q = 1 - p$. PVE was computed with the intention to allow for comparison of effect strength between timepoints, which have different sample sizes and different MAFs. In practice, this turns out to just be a monotonic transformation of the absolute posterior z -statistic $|\beta_p/\sigma_p|$, with more interpretable units.

3.2.9 Replication of eQTLs in a reference dataset

To validate the mega-analysis approach to eQTL mapping, I estimated the replication of significant eQTLs in a large independent reference. Due to the lack of large sample size eQTL maps specific to PBMC, I used the GTEx v8 whole blood dataset as my reference dataset ([54], $n = 670$, 51 % eGene rate). For lead variants called as significant at a given LFSR significance threshold in the HIRD dataset, for those variants that also exist in GTEx, I looked up their nominal p -values in GTEx. I then used `qvalue::qvalue_truncp` (v2.15.0*, implements theory from Storey *et al.* [309]) to estimate the proportion of those GTEx nominal p -values that are null (π_0), giving a measure of replication $\pi_1 = 1 - \pi_0$. The higher the π_1 , the higher the proportion of HIRD eQTLs at this significance threshold replicating in GTEx. However, the higher the significance threshold,

*<https://github.com/StoreyLab/qvalue>

the fewer variants will have p -values meeting this threshold in **HIRD**, and thus fewer GTEx p -values will be available for computing π_1 . More significant p -values in **HIRD** are also more likely to come from true **eQTLs** in general, so the higher the significance threshold, the lower the maximum nominal p -value from GTEx for those variants is likely to be. The π_1 procedure assumes a well-behaved p -value distribution with values over the full range $[0, 1]$, and reliability declines if the number of p -values is too small*, or the maximum p -value is much smaller than 1.

The mega-analysis had comparable replication rate to **RNA-seq**-only analysis for shared **eQTLs** at moderately stringent **LFSR** thresholds up to 10^{-5} , and better replication rate for very stringent **LFSR** thresholds (Fig. 3.7). This suggests the mega-analysis is not creating false positives due to technical effects from merging the expression data, and is preferred to either of the single-platform analyses. A caveat is this approach may be overestimating the replication rate as it does not take the direction or magnitude of **eQTL** effects into account. The numbers of **reQTLs** were too low to assess their replication using this method, and one might also not expect them to replicate in a baseline dataset such as GTEx whole blood, especially for those **reQTLs** significant only at post-vaccination timepoints.

3.2.10 Genotype interactions with cell type abundance

If the abundance of a particular cell type does truly modify the **eQTL** effect, then an interaction term between genotype and cell type abundance is required. As the additivity assumption no longer holds, a *ceteris paribus* interpretation does not make sense, as the effect of genotype holding cell type abundance constant depends on what value of cell type abundance you choose. One cannot adjust for modification just by including the main effect for cell type abundance; wrongly omitting a significant interaction term between cell abundance and genotype biases the estimation of the two main effects[†]. Given the modest sample size, I used a two-stage approach, where tests for interaction are only performed at a subset of tests. If the main effect estimates from main effect-only models (stage one) are used to filter **SNPs** for second stage testing, and are also independent from the interaction effect estimates in stage two, then the type I error can be controlled based on the number of interactions that are actually tested, rather than the number of interactions that could have been tested [68, 311]. It is unclear whether this assumption holds in practice, as being able to detect a main effect at least implies that gene is sufficiently expressed for **eQTL** mapping. Nevertheless, the two-stage approach is often used for **eQTL** mapping with an interaction term [68, 71, 85, 96]. As the main purpose of my interaction analyses was scanning for cell type modification at detected **reQTLs**, I chose to test for interactions only at the lead **eQTL** variant for each gene with a significant main **eQTL**, controlling the **FDR** with **BH**—an approach used by Peters *et al.* [68] and Kim-Hellmuth *et al.* [85].

Models with interactions between genotype and other predictors were fit using `lme4qt1` [312]. The model specification was identical to Eq. (3.4), except the addition of three interaction terms between genotype and each **xCell** score. Significance was assessed using the likelihood ratio test

*In <https://github.com/StoreyLab/qvalue/pull/6#commitcomment-26277751> the developers suggest “you usually need a few hundred p -values” to reliably compute π_1 .

[†]When a variable that is the function of another explanatory variable is omitted, this is known as functional form misspecification in the field of econometrics, a special case of omitted variable bias. Also see Mikucka *et al.* [310] for a review of bias caused by omitting significant interaction terms.

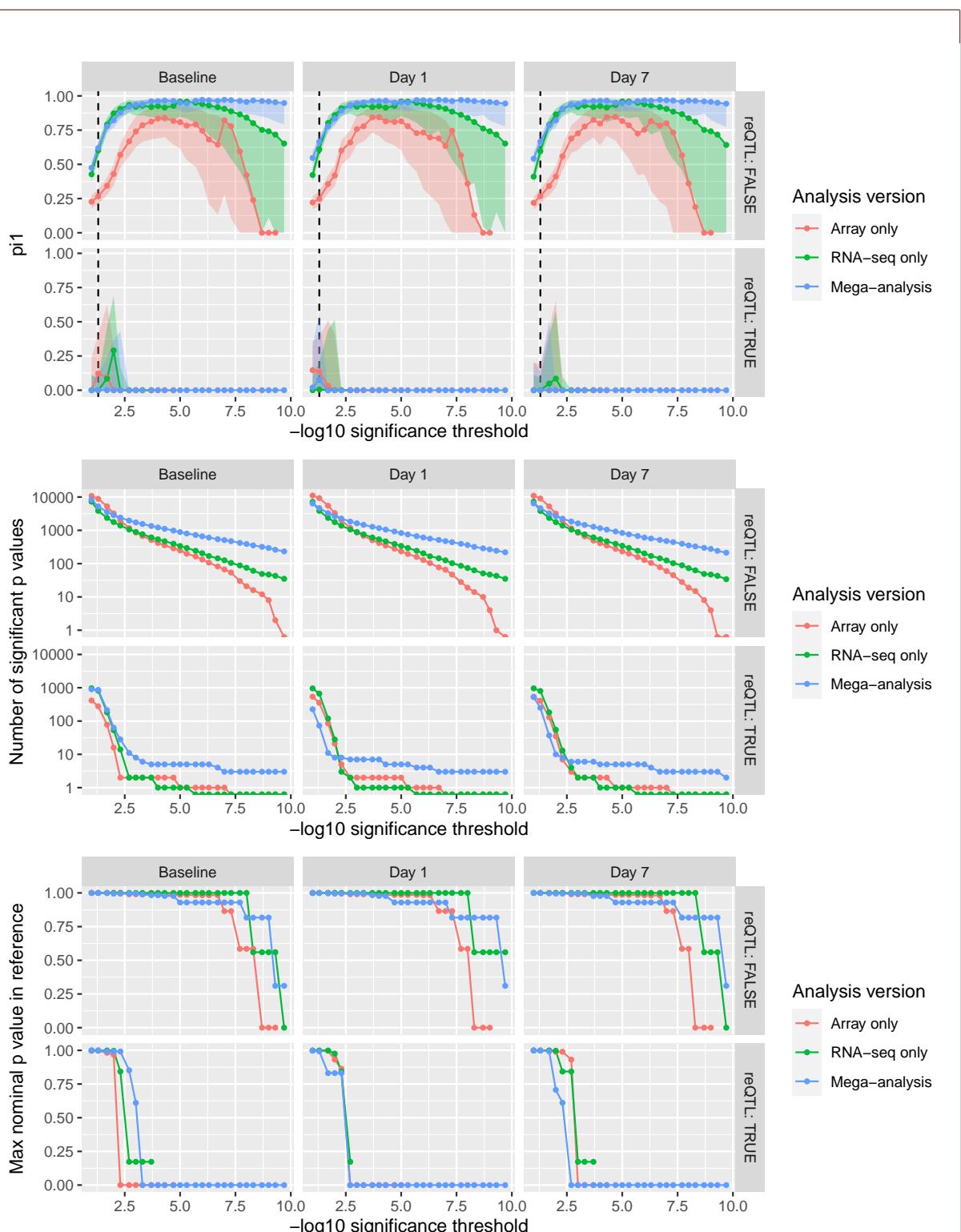


Figure 3.7: Replication rate π_1 of HIRD eQTLs in GTEx whole blood eQTL reference data. Three eQTL analyses were run in HIRD: array-only, RNA-seq-only, and a mega-analysis of the two datasets. The top panel shows π_1 replication in each analysis as a function of the significance threshold (LFSR), with the shaded region showing the 5th–95th percentile range of 1000 bootstraps. Vertical line shows LFSR = 0.05. The middle panel shows the number of significant HIRD eQTLs present in GTEx; this is the number of p -values available for computing π_1 . The computation is more reliable when there are ~ 1000 or more. The bottom panel shows the maximum nominal GTEx p -value for those variants used to compute π_1 . The computation is more reliable when the maximum is near one. Each panel is stratified into HIRD shared eQTLs and reQTLs.

(LRT) versus the nested model with no interaction terms. Note that although PEER factors are correlated with **xCell** scores (Fig. 3.5), Kim-Hellmuth *et al.* [74] also used this approach, claiming that the interaction term between genotype and **xCell** scores should still be interpretable. I attribute their claim to the fact that there are no interaction terms between genotype and PEER factors, so the coefficients for the genotype-**xCell** score interactions still have their standard interpretation: $\beta_x + \beta_{cx}c$ increase in \log_2 expression per unit of effect allele dosage, where β_x is the main effect of genotype, β_{cx} is the interaction effect, and c is **xCell** score.

3.2.11 Gene set enrichment analyses

Ranked gene set enrichment analyses with `tmod::tmodCERN0test` were conducted as described in Section 2.2.10, using blood transcription modules (BTMs) from Li *et al.* [240] (prefixed “LI”). Gene set overrepresentation analyses were run with `tmod::tmodHGtest` [241], which implements the hypergeometric test for enrichment in BTMs, controlling the FDR at 0.05 using the BH procedure. Gene set overrepresentation analyses were also run with `gprofiler2::gost` [313], which derives gene sets from Gene Ontology (GO), pathway databases (KEGG, Reactome, WikiPathways), regulatory motif databases (TRANSFAC, miRTarBase), protein databases (Human Protein Atlas, CORUM), and phenotype ontologies (HP). The default `gprofiler2` g:SCS method was used to control for multiple testing while accounting for the hierarchical structure of certain gene set databases like the GO. In both overrepresentation analyses, the 13 570 genes assayed by both array and RNA-seq were used as a custom background set.

3.2.12 Statistical colocalisation

Published GWAS and quantitative trait locus (QTL) summary statistics were downloaded for statistical colocalisation with per-timepoint HIRD eQTL summary statistics. Clinical blood count QTL maps generated by Astle *et al.* [303] in 173 480 European-ancestry participants were downloaded from ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2017-12-12/hematological_traits/. eQTL maps in fifteen FACS-sorted immune cell types generated by Schmiedel *et al.* [88] in a multi-ethnic cohort of 91 donors were downloaded from the eQTL Catalogue ([314], release 1, January 2020, <https://www.ebi.ac.uk/eqt1/>). The fifteen cell types included three naive innate immune cell types: classical monocytes ($CD14^{hi}CD16^-$), non-classical monocytes ($CD14^-CD16^+$), and NK cells; four naive adaptive immune cell types: B cells, $CD4^+$ T cells, $CD8^+$ T cells, and regulatory T cells (Treg); $CD4^+$ T cells and $CD8^+$ T cells stimulated with anti-CD3 anti-CD28 for 4 hours; and six $CD4^+$ memory T cell subsets: Th1, Th1/Th17, Th17, Th2, and memory Tregs. Inflammatory bowel disease (IBD) GWAS summary statistics generated by de Lange *et al.* [180] in a total of 59 957 European ancestry samples were downloaded from <https://www.ebi.ac.uk/gwas/studies/GCST004131>. Datasets were converted to GRCh38 coordinates with `rtracklayer::liftOver` (v1.46.0) [315] and harmonised to a standard format, matching variants between studies by genomic position and effect allele.

Multi-trait Bayesian colocalisation was performed using HyPrColoc [316]. HyPrColoc uses the pattern of per-variant summary statistics (betas and standard errors) from multiple traits in a locus to partition traits into clusters, where each cluster contains traits that share a causal variant. This can be seen as a multi-trait extension of pairwise Bayesian colocalisation methods

such as `coloc` [317]. Multi-trait colocalisation is more powerful than pairwise colocalisation for detecting causal variants shared between more than two traits, and large numbers of traits can be analysed simultaneously in a computationally efficient manner. The method formally assumes that studies generating the summary statistics for each trait are independent, but performs well even when there is complete sample overlap between traits [316]. If studies are non-independent, it is assumed the LD structure is the same across those studies (which holds in the case of multiple QTL maps generated from the same individuals). Each trait is assumed to have no more than one causal variant in the locus. Finally, it is assumed the causal variants for each trait are present in the input.

As with any Bayesian colocalisation method, the choice of priors and other algorithm parameters is influential. HyPrColoc implements variant-level priors where the prior depends on the number of traits a variant is causally associated with: `prior.1` is the prior probability that a variant is causal for one trait (default = 1×10^{-4}), and $1 - \text{prior.2}$ is the prior probability that a variant is causal for an additional trait, given it is causal for one trait (default = 0.98). The prior for a variant being causal for a third trait given it is causal for two traits is $1 - (\text{prior.2})^2$, and so on. In the two trait case, the setup is identical to `coloc` [317]. `prior.2` tends to be more influential than `prior.1`, as it controls the probability of association with more and more traits.

The posterior probability of colocalisation for a cluster of traits is the product of regional association and alignment probabilities. The regional association probability is the probability there is a shared association region within the locus for all the traits in the cluster, containing one or more causal variants. The alignment probability is the probability that regional association is due to a single causal variant, rather than one or more variants in strong LD. A branch and bound algorithm is run, starting with all traits in one cluster, then recursively partitioning traits into subsets, assessing regional association and alignment probabilities for subsets at each iteration. The end result is clusters of traits sharing a causal variant, with each cluster having a distinct causal variant. Only clusters with more than one trait, and regional association and alignment probabilities above `reg.thresh` (default = 0.5) and `align.thresh` (default = 0.5) are reported.

In sensitivity analyses using the `sensitivity.plot` function, I fixed the less influential `prior.1` at the default of 1×10^{-4} , then iterated over combinations of four choices of `prior.2` (0.98, 0.99, 0.995, 0.999), five choices of `reg.thresh` (0.5, 0.6, 0.7, 0.8, 0.9), and five choices of `align.thresh` (0.5, 0.6, 0.7, 0.8, 0.9). Each range starts at the default value and becomes more stringent, requiring stronger and stronger evidence for clusters of colocalised traits to be identified.

3.3 Results

3.3.1 Mapping reQTLs in the HIRD cohort

To characterise the effect of common host genetic variation on expression response to Pandemrix, I mapped *cis*-eQTLs for each gene (± 1 Mbp of the TSS) within each timepoint condition (baseline, day 1, and day 7), then conducted joint analysis of all three timepoints with `mashr` [281] to obtain per-timepoint posterior effect sizes (betas), posterior standard errors, and measures of

significance (the LFSR). At $\text{LFSR} < 0.05$, 6887/13 570 genes (50.8 %) were eGenes (genes with a significant eQTL) in at least one timepoint. The most significant tested variant over all timepoints was selected as the lead variant for each gene, then reQTLs were defined by comparing the effect size of this lead variant between each pair of timepoints. This guards against differences in effect size from differential tagging efficiency (of an assumed single causal variant), a potential issue if different variants are compared across timepoints. Fig. 3.8 shows patterns of sharing over timepoints for the lead variant for each of the 13 570 genes, illustrating the difference between calling reQTLs using a significance threshold and a difference in betas method. For example, there were 85 eQTL-eGene pairs significant only at day 1 post-vaccination ($\text{LFSR} < 0.05$); of these, only 40/85 were called as reQTLs by the difference in betas method. The difference in betas method is more strict because calling by significance alone would call a reQTL for an eQTL with $\text{LFSR} = 0.049$ at baseline and $\text{LFSR} = 0.051$ at day 1, even if the effect sizes are similar. Shared eQTLs were well-replicated in GTEx whole blood (Fig. 3.7).

The largest number of eGenes was detected at baseline, reflecting the larger sample size compared to other timepoints. Most eQTLs were shared across timepoints; these were also the strongest eQTLs in terms of both maximum absolute beta and PVE across timepoints, highlighting the power advantage for mapping shared effects granted by joint analysis. Based on difference in effect size between any pair of timepoints (nominal $p < 0.05$), 1154/6887 (16.8 %) eQTLs were classified as reQTLs. Of these, 690/1154 were reQTLs between both day 1 vs. baseline and day 7 vs. baseline, and only 23/1154 were unique to the day 7 vs. day 1 comparison, indicating most reQTL effects were differences between pre- and post-vaccination (Fig. 3.9).

3.3.2 Characterising reQTLs post-vaccination

To characterise the eGenes associated with post-vaccination reQTLs, I ranked eGenes by the increase in PVE for their associated reQTLs from baseline to day 1 and baseline to day 7, then performed ranked gene set enrichments with `tmod::tmodCERN0test` [241]. The same four modules were significant at both post-vaccination timepoints: “immune activation - generic cluster” (LI.M37.0, day 1 FDR = 1.28×10^{-6} , day 7 FDR = 3.39×10^{-6}), “enriched in monocytes (II)” (LI.M11.0, day 1 FDR = 4.69×10^{-3} , day 7 FDR = 1.88×10^{-2}), “cytoskeleton/actin (SRF transcription targets)” (LI.M145.0, day 1 FDR = 2.07×10^{-2} , day 7 FDR = 2.04×10^{-2}), and “MHC-TLR7-TLR8 cluster” (LI.M146, day 1 FDR = 2.07×10^{-2} , day 7 FDR = 2.04×10^{-2}). These enrichments are weak, but consistent with immune activation driving post-vaccination reQTLs. Given that TLR7 and TLR8 are primarily expressed in monocytes, macrophages, and DCs [318], and SRF is a regulator of the cytoskeleton in macrophages [319], there is suggestive evidence reQTLs may be enriched in genes specific to these phagocytotic APCs.

Changes in PVE do not capture changes in allelic direction. I classified post-vaccination reQTLs into one of three effect types: magnified, where the beta increases after vaccination but remains the same sign; dampened, where the beta decreases after vaccination but remains the same sign; and opposite, where the allelic direction changes after vaccination. As LFSR quantifies uncertainty in the sign of the effect, I did not make this classification for reQTLs that were not significant both at baseline and post-vaccination—the effect type for these are unclear. The classifications are shown in Fig. 3.10, plotting all 6887 shared or reQTLs by their

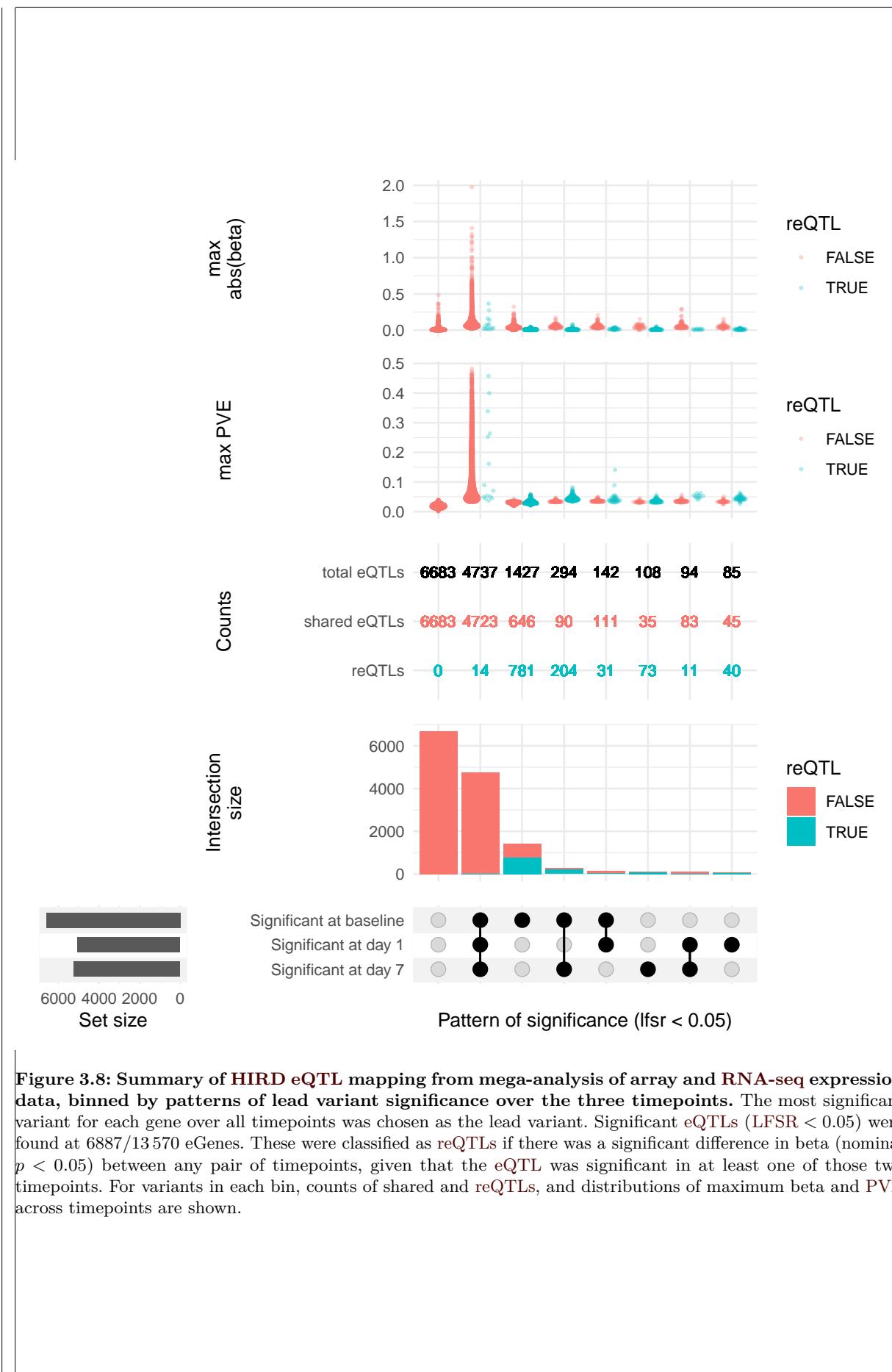


Figure 3.8: Summary of HIRD eQTL mapping from mega-analysis of array and RNA-seq expression data, binned by patterns of lead variant significance over the three timepoints. The most significant variant for each gene over all timepoints was chosen as the lead variant. Significant eQTLs ($LFSR < 0.05$) were found at 6887/13 570 eGenes. These were classified as reQTLs if there was a significant difference in beta (nominal $p < 0.05$) between any pair of timepoints, given that the eQTL was significant in at least one of those two timepoints. For variants in each bin, counts of shared and reQTLs, and distributions of maximum beta and PVE across timepoints are shown.

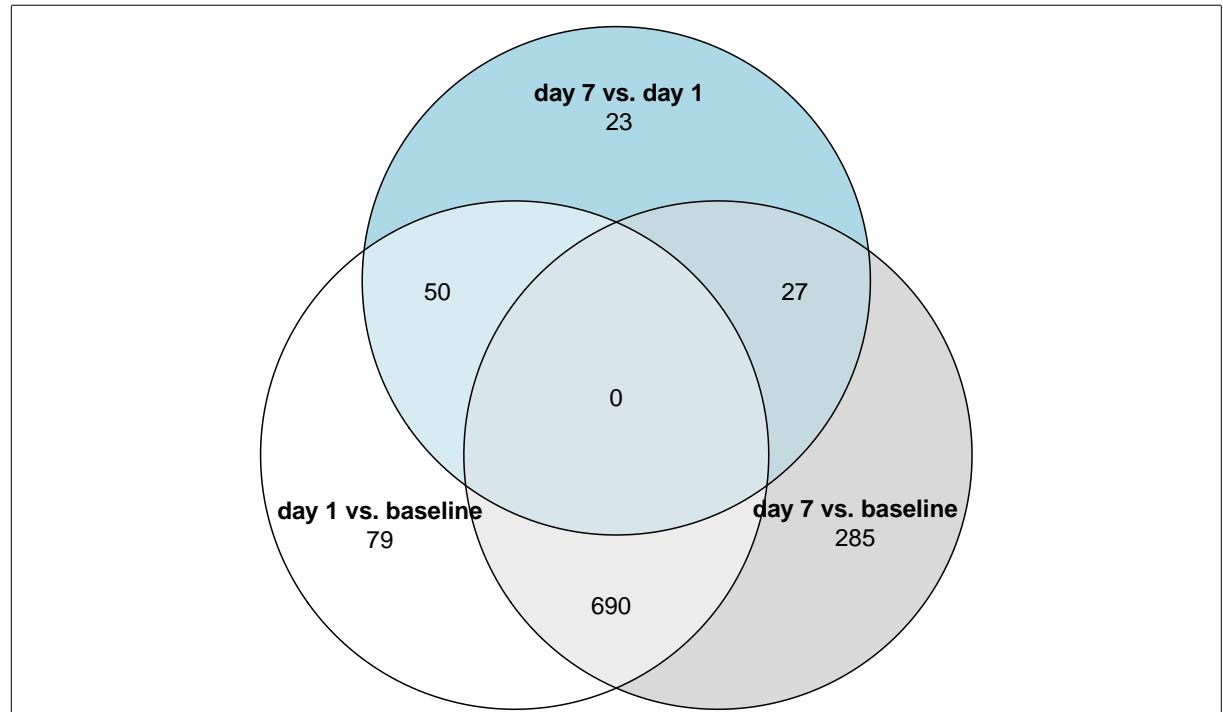


Figure 3.9: eGenes where the lead eQTL was a reQTL between a pair of timepoints. reQTLs were observed for 1154 unique eGenes, defined by a significant difference in beta between timepoints (nominal $p < 0.05$).

distance relative to the eGene TSS. Shared eQTLs have difference in beta z -statistics close to zero, and are concentrated close to the TSS as expected. reQTLs have a distribution of mostly negative z -statistics clearly separated from the shared eQTLs at both day 1 and 7, and these are mostly unclear or opposite rather than dampened effect types. Many of these unclear effects may actually be dampening, but as the sample size is greatest at baseline, dampening effects are hard to distinguish from drops in power at post-vaccination timepoints, whereas an opposite effect significant in both timepoints is unambiguous.

Fig. 3.10 also shows that reQTLs tended to be distributed evenly across the entire *cis* window, raising the question of whether they are enriched in false positives. A nominal $p < 0.05$ difference in betas threshold—although stricter than many other methods (see Section 3.2.8)—may still be too lax for calling reQTLs, so I applied a stronger BH FDR = 0.2 threshold. At this threshold, the only remaining reQTL was at day 1 was for *ADCY3* (nominal $p = 8.68 \times 10^{-6}$, FDR = 0.12)—the next smallest FDR value was 0.65. At day 7, 676 significant reQTLs had FDR < 0.2, of which 221 were opposite effects. Performing gene set overrepresentation analysis on the set of 221 eGenes to identify a shared biological signature was relatively uninformative, and revealed only one enrichment for genes *PRKACB*, *PRKACA*, *SAR1B*, and *APOE* in “Plasma lipoprotein assembly” (Reactome pathway identifier R-HSA-8963898, set size = 11, adj. $p = 0.01$). Since Fairfax *et al.* [59] found cases of opposite reQTL effects between B cells and monocytes, and B plasma cells but not monocytes were increased in abundance at HIRD day 7 [162], I also performed gene set overrepresentation analysis using BTMs to detect if eGenes related to B cells were enriched at day 7. No significant enrichments were identified.

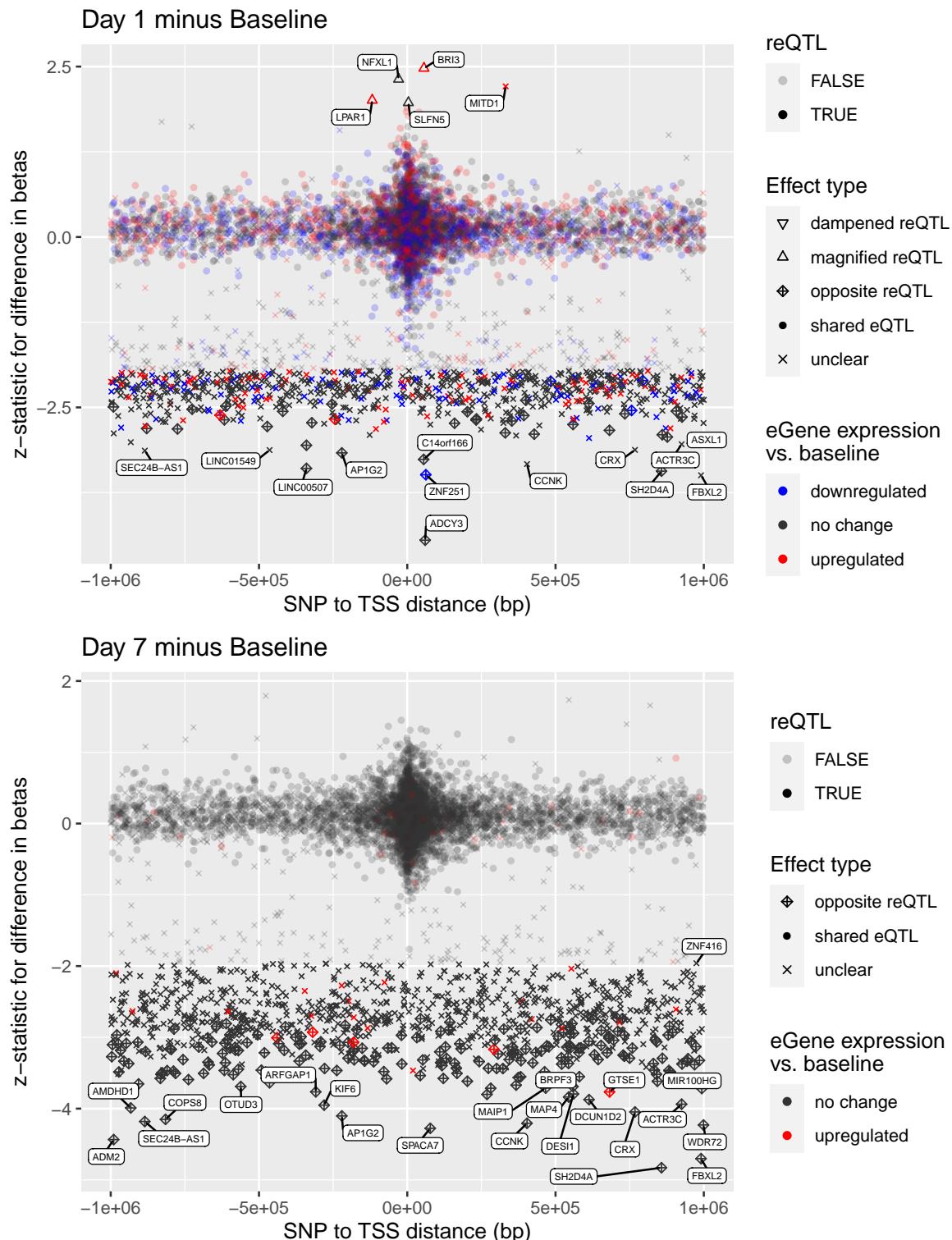


Figure 3.10: *z*-statistic for difference in beta post-vaccination versus baseline for shared and reQTLs, against distance from the eGene TSS. For each plot, all eQTLs significant in either timepoint are shown. Shared eQTLs can only have the shared effect type. An unclear effect type indicates the eQTL in question is not significant in both timepoints. Allelic direction of effect is aligned so that the beta at baseline is positive.

3.3.3 Exploring possible mechanisms generating reQTLs

3.3.3.1 Differential expression of genes with reQTLs

As gene set analyses based on the effect sizes of reQTLs at different timepoints had been largely uninformative, I considered whether reQTLs could be characterised by shared mechanisms. One mechanism that could generate reQTLs is differential expression, where an eQTL is not detected at baseline because the eGene is not expressed, and vaccine-stimulated upregulation reveals the effect post-vaccination.

Fig. 3.10 shows whether each eGene was up or downregulated at the timepoint based on the DGE analyses in Chapter 2. Visually, a large number of reQTLs occur without corresponding differential expression. Statistically, compared to genes without reQTLs, genes with reQTLs were less likely to be differentially expressed at day 1 post-vaccination (26.5 % for genes with reQTLs, 42.3 % for genes without reQTLs, Fisher's test $p < 2.20 \times 10^{-16}$). This was also the case when restricting the scope to only eGenes (26.5 % for genes with reQTLs, 47.9 % for genes with shared eQTLs, Fisher's test $p < 2.20 \times 10^{-16}$). At day 7, no significant difference was observed comparing genes with and without a reQTL (2.2 % for genes with reQTLs, 1.4 % for genes without reQTLs, Fisher's test $p = 0.05$), but compared to genes with shared eQTLs, genes with reQTLs were more likely to be upregulated (2.2 % for genes with reQTL, 1.1 % for genes with shared eQTLs, Fisher's test $p = 0.01$). Twenty-two genes with both day 7 reQTLs and upregulated expression were strongly enriched within gene sets related to the cell cycle, such as “mitotic cell cycle” (GO biological process term GO:0000278, term size = 914, intersection size = 12, `gprofiler2::gost` [313] adj. $p = 1.42 \times 10^{-4}$), “cell cycle (I)” (LI.M4.1, module size = 137, intersection size = 12, `tmodHGtest` [241] FDR = 1.41×10^{-16}), and “mitotic cell cycle in stimulated CD4 T cells” (LI.M4.5, module size = 33, intersection size = 3, `tmodHGtest` FDR = 1.35×10^{-3}). However, these 22 genes previously appeared in Fig. 3.10 as having reQTL with decreased or opposite effect at day 7 versus baseline, making it implausible that the generating mechanism is increased detection power due to upregulation. The enrichment for cell cycle is likely driven by the DGE signal alone, especially as similar cell cycle gene modules were detected to be strongly upregulated at day 7 in Section 2.3.1.2.

The presence of reQTLs without DGE is exemplified by the strongest reQTL at each day. The only significant day 1 reQTL at difference in betas FDR < 0.2 was for *ADCY3* (Fig. 3.11). Computing the PVEs, this reQTL explained 1.9 % of *ADCY3* expression variation at baseline, increasing to 14.1 % at day 1, yet *ADCY3* was not differentially expressed from baseline to day 1 ($\log_2 FC = 0.10$, LFSR = 0.26). The strongest day 7 reQTL was at *SH2D4A* (difference in betas FDR = 0.02, Fig. 3.12). Here, the reQTL variant explained similar amounts of expression variation at baseline (PVE = 8.2 %) and day 7 (PVE = 9.0 %), with opposite directions of effect. Again, there was no differential expression. There is strong evidence that many post-vaccination reQTLs are generated by mechanisms unrelated to DGE.

3.3.3.2 Genotype by cell type abundance interaction effects

The presence of cell type-specific eQTL effects combined with changes in cell abundance between timepoints was considered as an alternate explanation generating reQTLs. Even if an eGene is

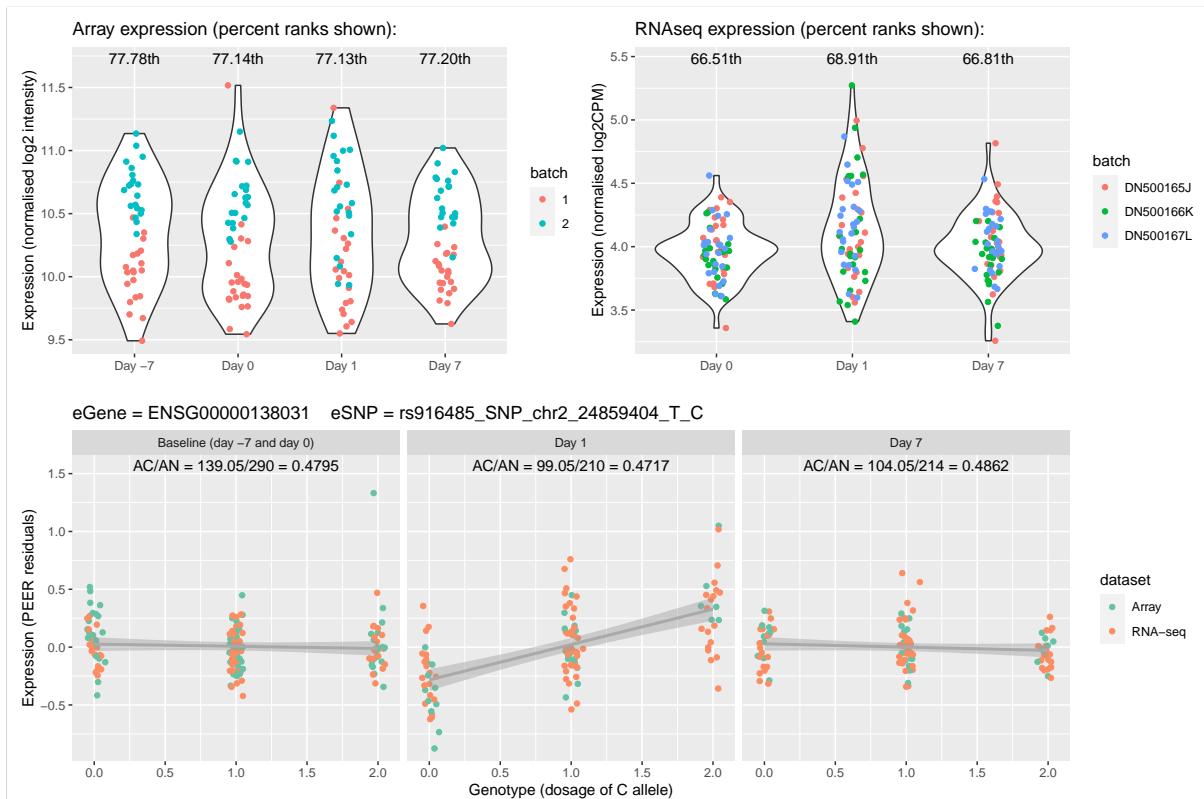


Figure 3.11: Expression and lead eQTL of *ADCY3* over study timepoints. Normalised array (top-left) and RNA-seq (top-right) expression before batch effect correction with ComBat and eQTL mapping. Bottom: eQTL effects at each timepoint condition in the mega-analysis of array and RNA-seq data.

not differentially expressed on average in bulk expression data, the composition of cell types that are the source of that gene's transcripts can change. xCell [298] enrichment scores were used to estimate abundance of seven PBMC cell types from the expression data. After pruning highly correlated cell types to avoid multicollinearity, standardised scores for monocytes, NK cells, and plasma cells were tested for interaction with genotype. Within each timepoint, full eQTL models including genotype main effect, the three cell type abundance main effects, and three cell type-genotype interaction terms, were fit using `lme4qt1` [312], then compared to a nested model excluding the three interaction terms with a LRT. Significant cell type interactions were detected at 16/1154 reQTL-gene pairs in at least one timepoint (BH FDR < 0.05). Fifteen were significant in only one timepoint: baseline (*SLAMF8*, *CSE1L*, *MAST1*, *DLGAP1*), day 1 (*ZNF519*, *LPAR1*, *ADCY3*, *NAA20*, *EPB41L5*), or day 7 (*APOL6*, *ADAR*, *ADAM17*, *UHRF2*, *MST1*, *CUL1*).

For *ADCY3* at day 1 (full vs. nested FDR = 9.54×10^{-5}), although the genotype effect was 0.26 (standard error = 0.03) in the nested model; the estimate in the full model was -0.01 (0.07), with the three cell type-genotype interaction term estimates being 0.21 (0.05) for monocytes, -0.01 (0.04) for NK cells, and 0.02 (0.07) for plasma cells. The small magnitude of the genotype main effect in the full model compared to the nested model suggests the eQTL effect is driven largely by the monocyte score (or a cell type that is highly correlated with monocyte score in Fig. 3.2). In the case where the monocyte score is zero (representing an average abundance across samples, as scores were standardised), the effect of increasing genotype dosage on *ADCY3*

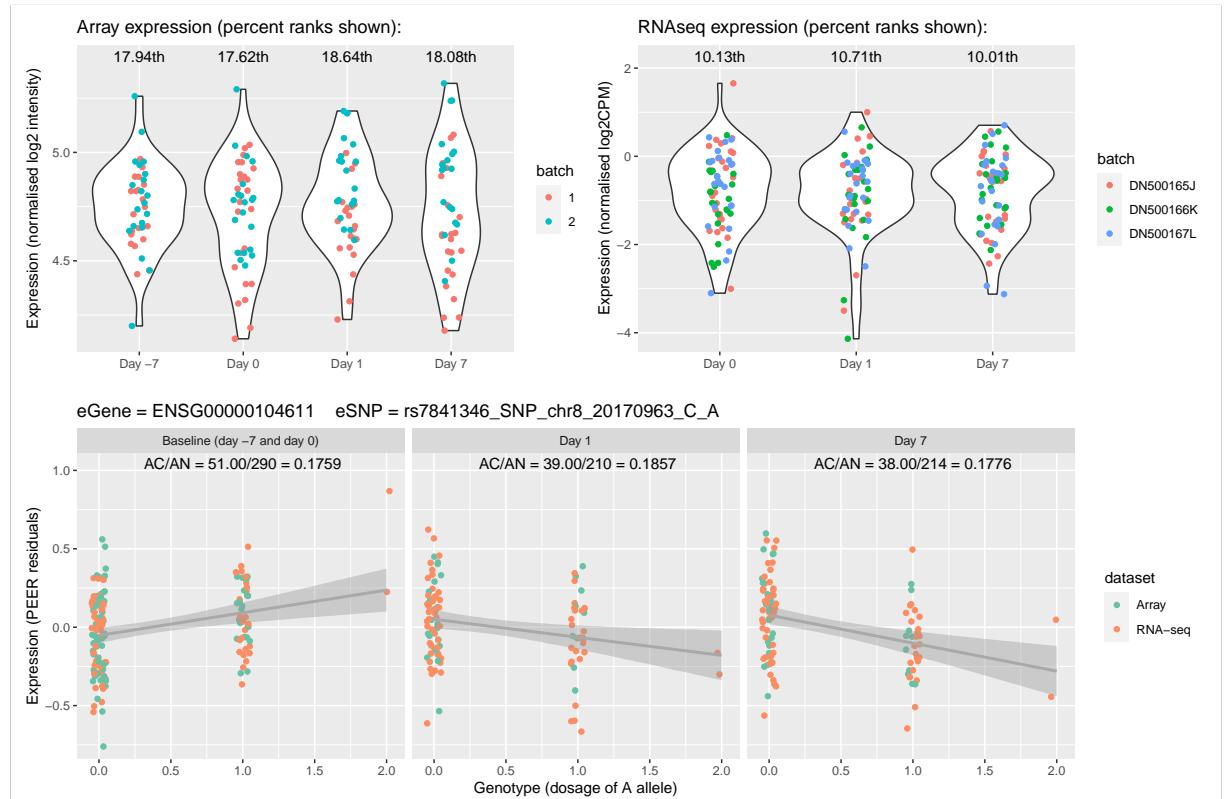


Figure 3.12: Expression and lead eQTL of *SH2D4A* over study timepoints. Normalised array (top-left) and RNA-seq (top-right) expression before batch effect correction with ComBat. Bottom: eQTL effects at each timepoint condition in the mega-analysis of array and RNA-seq data.

expression is minimal. Fig. 3.13 and 3.14 illustrate this effect. Monocyte abundance has no effect on expression at baseline, increases after vaccination, and modifies the effect of genotype on expression at day 1. It is feasible that the mechanism generating reQTLs at the remainder of these genes also involve cell type-specific eQTL effects, but unlike at *ADCY3*, I have not yet examined which of the three cell abundance scores have the greatest contributions.

3.3.3.3 Colocalisation with external QTL datasets at the *ADCY3* locus

The day 1 *ADCY3* reQTL is of particular interest, as reQTLs were also found for *ADCY3* in blood approximately 1 day after stimulation with TIV [94], rhinovirus [83], and *Mycobacterium leprae* [92]. The locus containing *ADCY3* has also been implicated in disease risk for immune-mediated inflammatory diseases (IMIDs) such as IBD [180], and *ADCY3* expression in immune cells in gut mucosa has been suggested to contribute to Crohn's disease (CD) risk (a subtype of IBD) [320]. Aside from monocytes, *ADCY3* is expressed in a wide range of immune cells; CD4⁺ T cells, CD8⁺ T cells, B cells, and NK cells (Fig. 3.15). Identifying cell type-dependent eQTLs through genotype-cell type abundance interaction terms cannot distinguish between cell types with highly correlated abundances [74]; the similar contributions to xCell score PC1 by monocyte, CD4⁺ T cell, and CD8⁺ T cell scores indicates cell types with *ADCY3* expression are indeed correlated in HIRD (Fig. 3.3). Given the *ADCY3* locus is associated with response to a wide range of immune stimuli, and also an IMID, I conducted colocalisation analysis to test if shared causal variants may be driving these associations, and to determine which among the

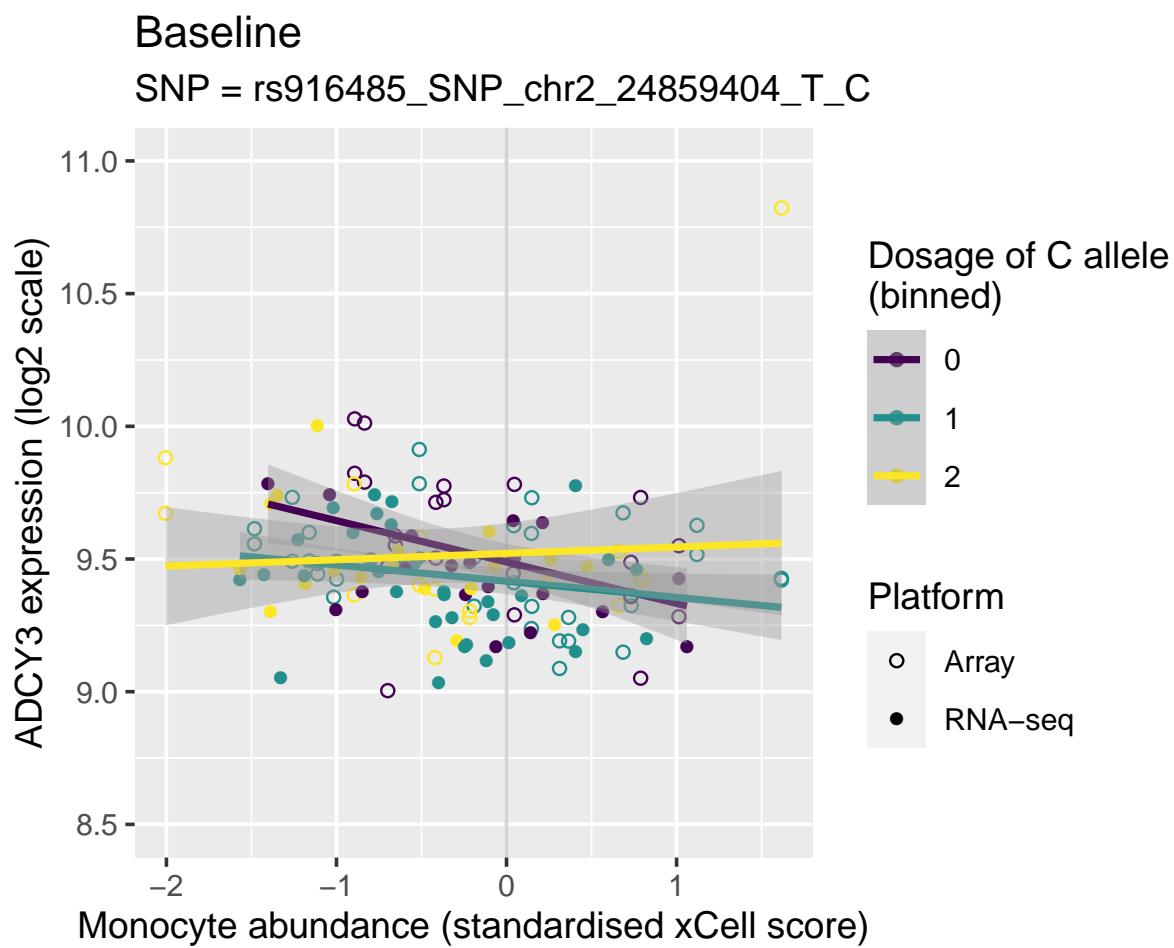
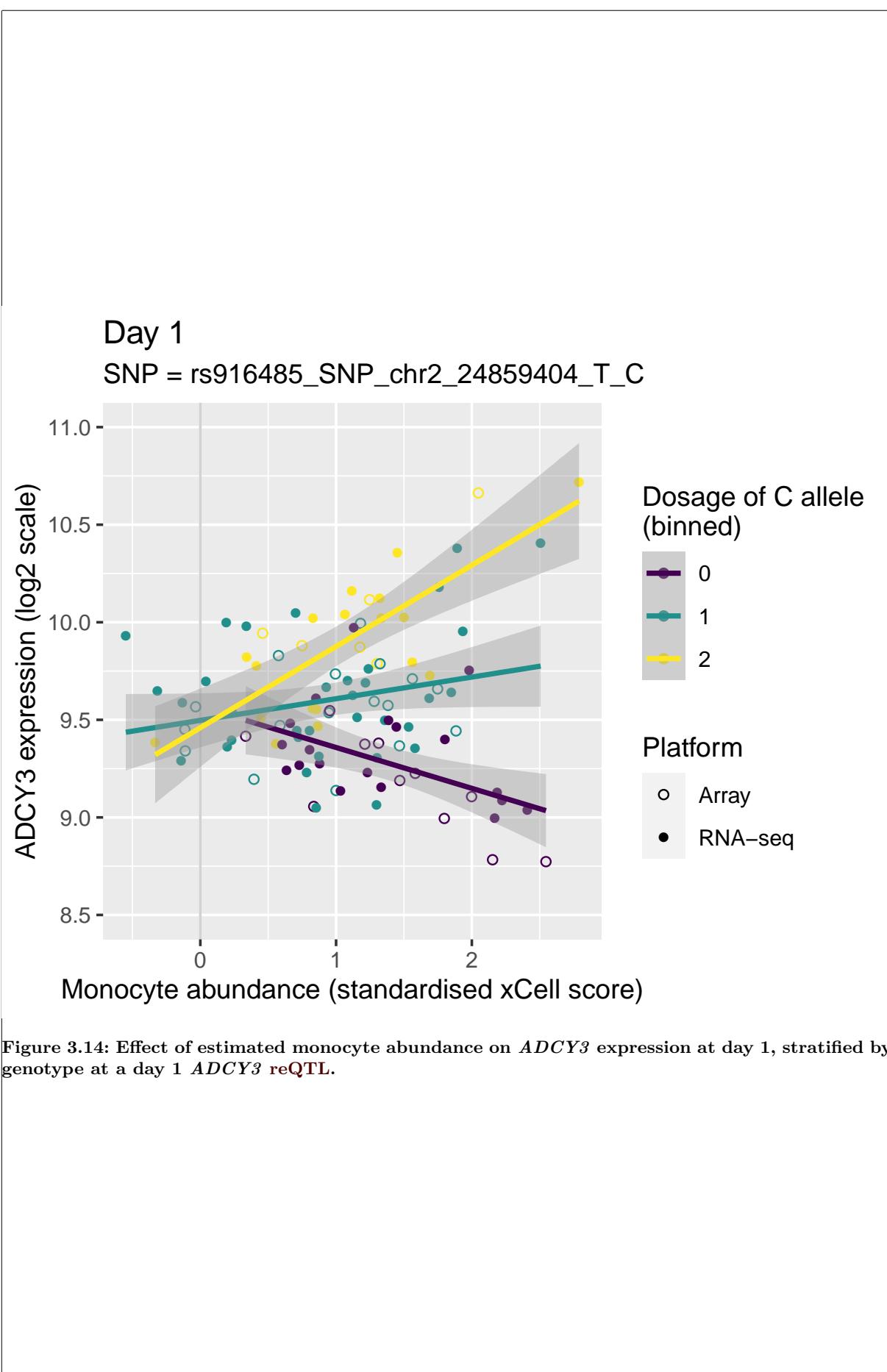


Figure 3.13: Effect of estimated monocyte abundance on *ADCY3* expression at baseline, stratified by genotype at a day 1 *ADCY3* reQTL.



correlated cell types expressing *ADCY3* are mostly likely responsible.

In a ± 500 Mbp window around the lead reQTL variant rs916485, I performed Bayesian multi-trait colocalisation (HyPrColoc [316]) of the three per-timepoint *ADCY3* eQTL summary statistics with external summary statistics: *ADCY3* eQTL in 15 sorted immune cell populations from Schmiedel *et al.* [88], monocyte count QTL from Astle *et al.* [303], and IBD GWAS from de Lange *et al.* [180]. There were 1054 variants present in all 20 sets of summary statistics.

HyPrColoc identifies clusters of traits that colocalise at different causal variants in the locus. As Bayesian colocalisation can be sensitive to the choice of priors, I performed a sensitivity analysis iterating over configurations of priors and other algorithm parameters, ranging from default to more stringent parameter values. Two stable clusters were identified across 100 configurations of parameters (Fig. 3.16). A set of three traits—*ADCY3* expression at HIRD day 1, and in naive classical and non-classical monocytes—clustered in $\sim 65\%$ of tested configurations. A set of nine traits—IBD, and expression in eight naive and memory CD4 $^{+}$ T cell subsets—clustered in $\sim 90\%$ of tested configurations. The remaining traits did not robustly cluster with any other traits over the tested configurations, except for the rare inclusion of HIRD baseline *ADCY3* expression into the larger cluster for less stringent configurations. The value of prior.2 (the probability that a variant associated with at least one trait is not associated with any additional traits) was subsequently set to 0.99 (default = 0.98) to increase stringency, preventing this inclusion of baseline HIRD expression into the larger cluster. The values of other priors and algorithm parameters were left at their defaults, producing the final clustering shown in Fig. 3.17.

Under the final configuration, the posterior probability that all traits in the cluster share a causal variant was 0.94 for the smaller cluster (HIRD day 1 and monocyte expression), and 0.98 for the larger cluster (CD4 $^{+}$ T cell expression and IBD). Distinct candidate causal variants were proposed for each cluster: for the smaller cluster, rs7567997, an intronic variant 45 kbp downstream of the canonical *ADCY3* TSS; and for the larger cluster, rs713586, a variant 15 kbp upstream of the TSS. In both cases, the variant explained all of the posterior probability for the cluster, but as the analysis was restricted to the 1054 variants present in all datasets, there is ample chance the true causal variants were not included. When it comes to fine mapping, it would be more appropriate to perform it using the dataset with the densest genotyping in each cluster. Nevertheless, the two main clusters being distinct from one another, and from non-colocalising traits across many configurations, still supports the existence of distinct causal variants, even if they may be unobserved. For HIRD day 1 expression of *ADCY3*, the more relevant cell type appears to be monocytes, not a correlated cell type like CD4 $^{+}$ T cells—and vice versa for IBD. The clustering was robust despite the data containing no stimulated monocyte subsets. This eQTL effect is readily observable at baseline, and appears to be more significant in naive classical than non-classical monocytes in the Schmiedel *et al.* [88] data (Fig. 3.17). No colocalisation with blood monocyte count was observed, so the reQTL does not appear to affect monocyte abundance in general. I believe a variant that affects ability to increase monocyte counts post-vaccination can also be ruled out, as in that case the effect of genotype on expression is entirely mediated through the effect of genotype on monocyte abundance, so having cell abundance scores as covariates in the regression should eliminate that effect. Thus I hypothesise that a plausible mechanism generating the day 1 reQTL signal in HIRD is an increase in abundance of (classical)

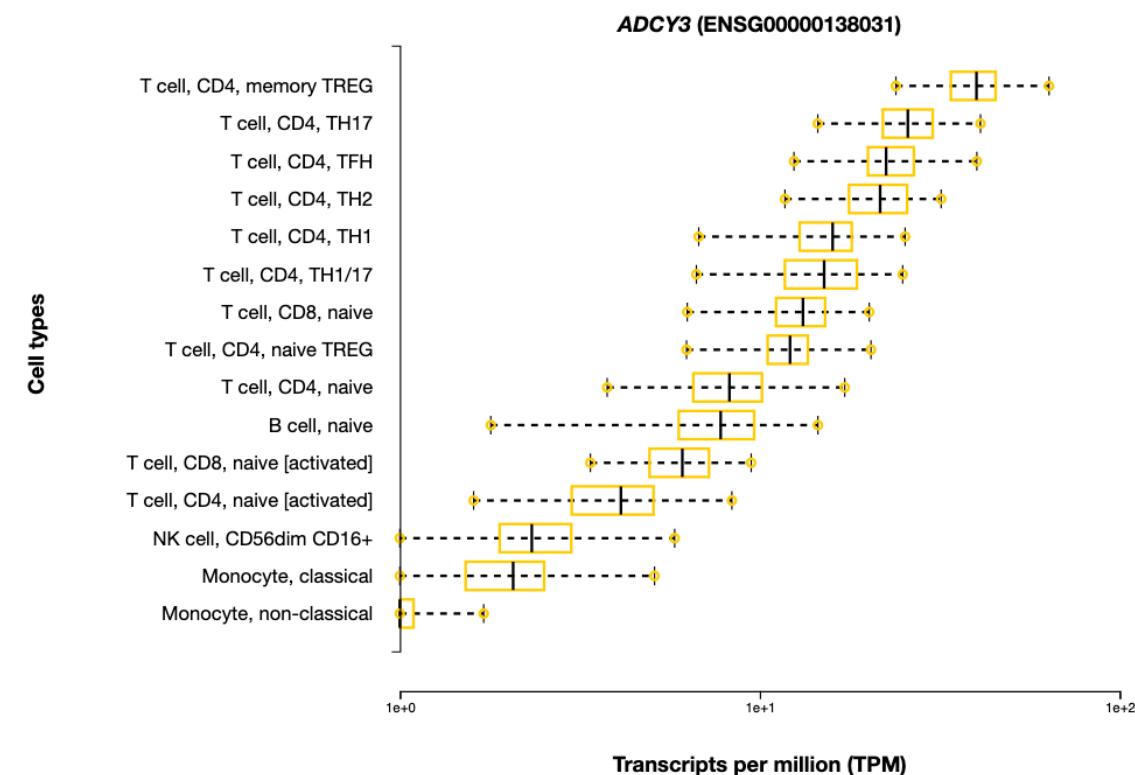
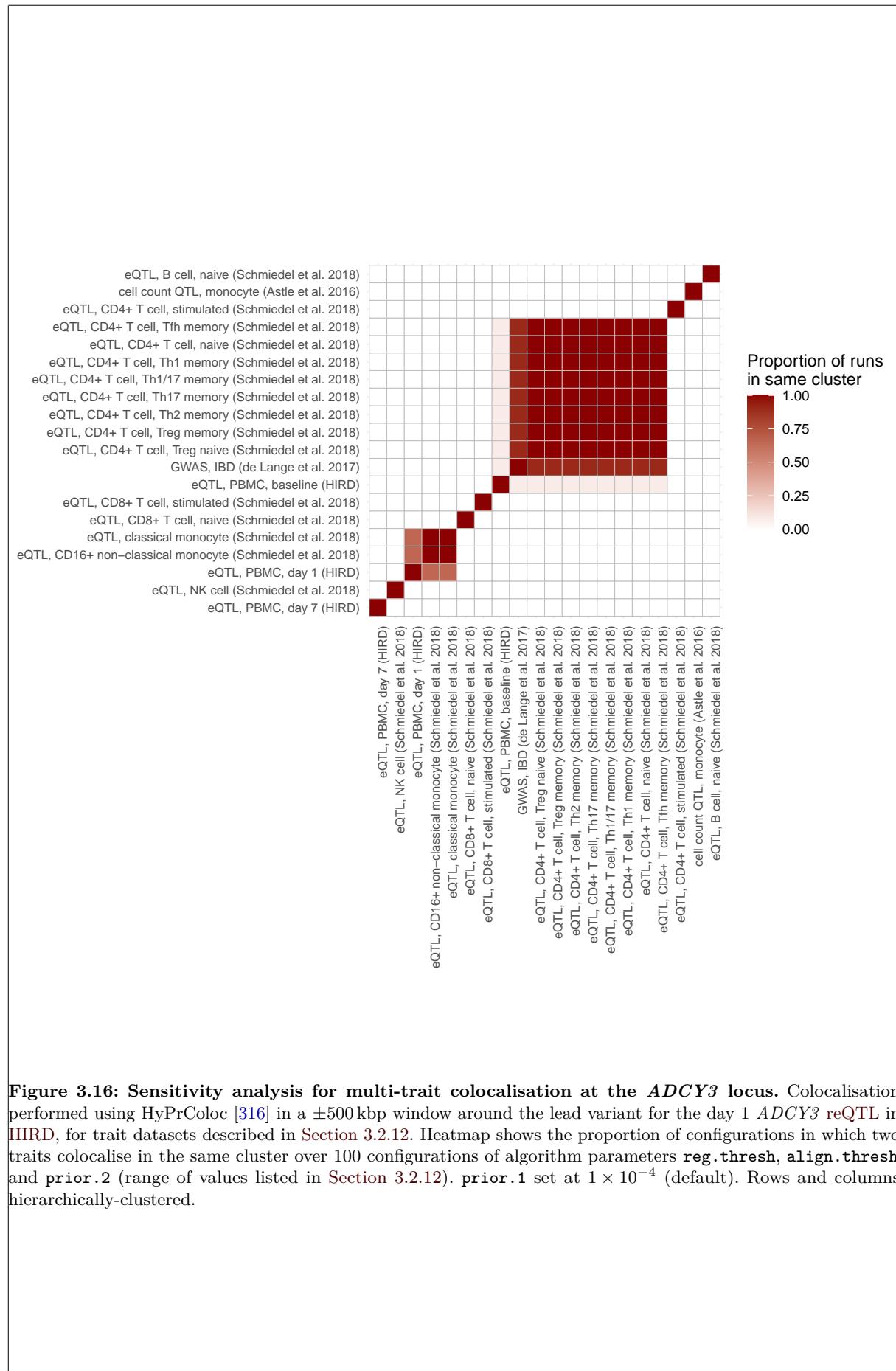


Figure 3.15: Expression of *ADCY3* in sorted immune cell subsets. Figure from Schmiedel *et al.* [88]. DICE (database of immune cell expression, expression quantitative trait loci, and epigenomics), <https://dice-database.org/genes/ADCY3>, accessed Nov 2020.

monocytes at day 1 post-vaccination, increasing the proportion of *ADCY3* transcripts in the bulk data originating from monocytes, thus making an eQTL specific to monocytes—not just stimulated monocytes—more readily detectable. This is the scenario where monocyte abundance modifies the effect of genotype on expression.



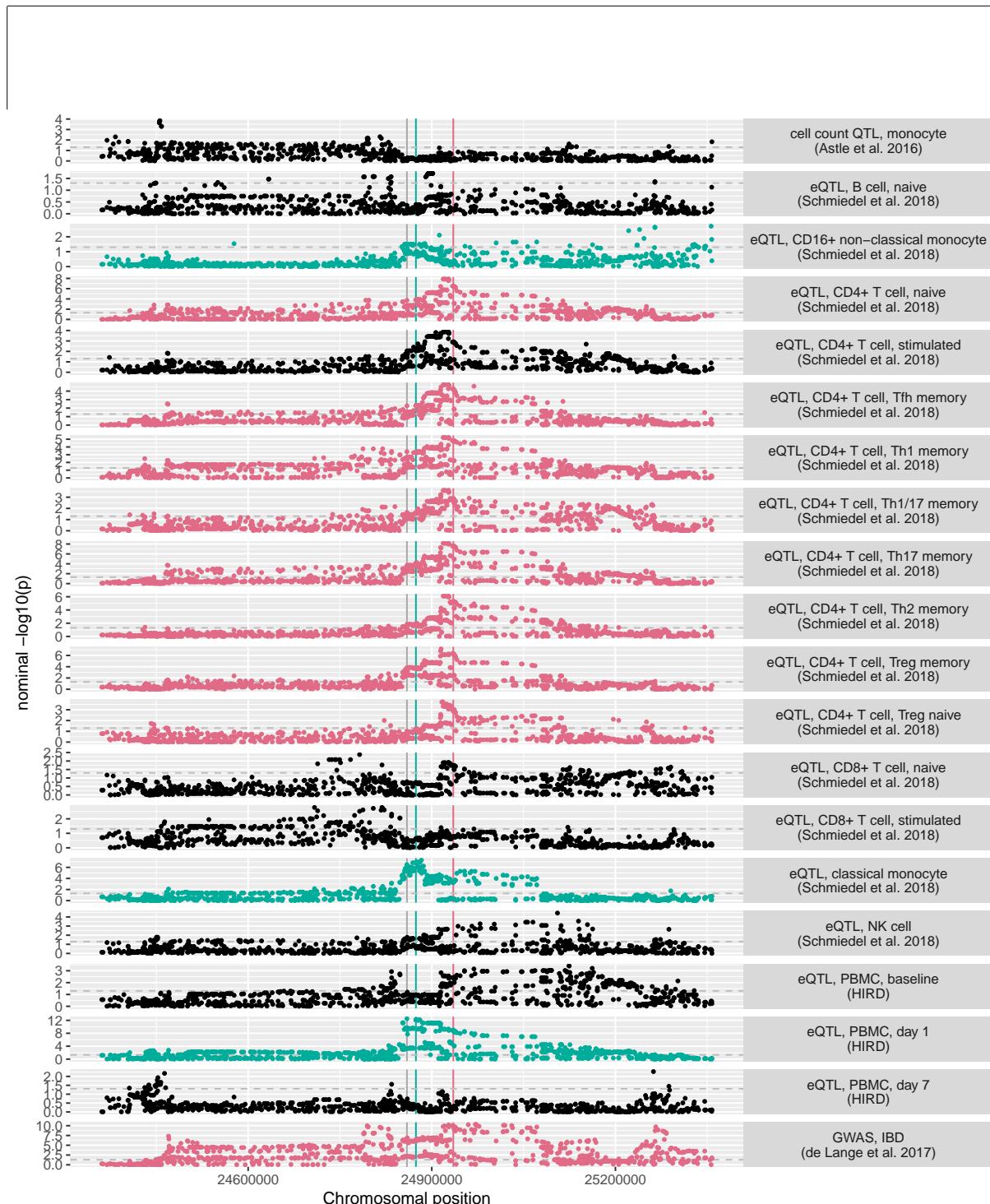


Figure 3.17: Multi-trait colocalisation at the *ADCY3* locus. Colocalisation performed using HyPrColoc [316] in a ± 500 kbp window around the lead variant for the day 1 *ADCY3* reQTL in HIRD (vertical grey line). Traits are monocyte cell count (Astle *et al.* [303]), *ADCY3* expression in sorted immune cell subsets (Schmiedel *et al.* [88]), *ADCY3* expression at HIRD timepoints, and IBD (de Lange *et al.* [180]). Locus plots show summary statistics for 1054 variants present in all datasets. Traits in red and green represent two distinct clusters each hypothesised to be driven by a shared causal variant (vertical red and green lines). Non-colocalising traits are shown in black. Horizontal dashed lines show nominal $p = 0.05$. Default values for priors and algorithm parameters used, except `prior.2 = 0.99`.

3.4 Discussion

Just as Pandemrix vaccination was found to induce extensive changes in the transcriptome in Chapter 2, it also induces changes in the regulatory architecture of gene expression. In a mega-analysis of array and RNA-seq datasets, *cis*-eQTL were detected for 50.8 % (6887/13 570) of genes in at least one timepoint, with the majority replicating in the much larger GTEx whole blood dataset. This is a substantial eGene rate given the modest per-timepoint sample size in HIRD, reflecting the gain in effective sample size from joint mapping over multiple conditions. Defining reQTLs by a significant difference in beta of the same eQTL between two timepoints, 1154/6887 (16.8 %) of lead eQTLs were classified as reQTLs. This is comparable to estimates of a 3–18 % reQTL rate between monocytes in different stimulation conditions by Kim-Hellmuth *et al.* [85], who also used a beta comparison method. The method is relatively stringent for calling reQTLs, avoiding both threshold effects where significant and non-significant eQTLs may have very similar betas, and discovery power biases caused by sample size differences between conditions. Indeed, had reQTLs been called by significance alone, 1427 reQTLs would have been detected with effects specific to baseline, the timepoint with the largest sample size in HIRD. There is growing consensus in the literature that most eQTLs are shared between conditions such as tissue and cell-type [74, 281, 321, 322], and that high estimates of >50 % condition-specificity based on significance thresholds (e.g. [276]) are overestimated. A counter-argument is that many studies overestimate sharing by calling condition-specific effects in LD as shared [322]. Here I compared the same gene-tag SNP pair across timepoints, but distinct causal, condition-specific variants may be tagged in such a way that the effect size of the tag SNP ends up similar—multi-trait colocalisation would be required to truly confirm a shared eQTL.

Gene set enrichment analyses to identify shared biological processes among target genes for reQTLs were generally uninformative. Genes targeted by reQTLs that explained more variation in expression post-vaccination were enriched for immune activation, with weaker enrichments related to APCs. This misses the full picture, as many of the strongest reQTLs were those with opposite sign effects at baseline and post-vaccination, but little change in PVE. Prevalence of opposite sign effects between pairs of conditions has been previously described in multi-tissue studies: in Fu *et al.* [76], the proportion of opposite sign effects among all reQTLs between five tissues was 4.4 %. In HIRD, I found an unexpectedly high proportion: 39/819 (4.8 %) for day 1 reQTLs, and 211/1002 (21.1 %) for day 7 reQTLs (Fig. 3.10). Given the global change in expression versus baseline was larger at day 1 than at day 7 (as described in Chapter 2), the larger number of strong reQTLs at day 7 was also unexpected. The genes with these opposite sign effects were not significantly enriched in any of the gene sets or BTMs I tested.

Post-vaccination DGE was considered as a mechanism that might generate reQTLs. As in Kim-Hellmuth *et al.* [85] and Davenport *et al.* [96], the overlap between differentially expressed genes and genes with reQTLs in HIRD was poor. Only at day 7 were genes with reQTLs more likely to be differentially expressed than genes without reQTLs—specifically after excluding genes without an eQTL from the analysis. The genes with both day 7 reQTLs and day 7 upregulation were enriched in cell cycle GO terms, but it is unclear how this may lead to generation of opposite sign effects, and the enrichment may largely have been driven by the DGE signal rather than the reQTL one. As described in Section 3.1.2, to define genes important to TIV response, Franco

et al. [94] made heavy use of the overlap of genes with DGE and reQTLs, followed by gene set enrichment. Unfortunately, their filtering before enrichment selected genes with either DGE or reQTL, making it difficult to assess which criteria contributed more to the significant enrichments they observed in the antigen presentation pathway. As noted by Davenport *et al.* [96] and Cuomo *et al.* [323], it may be that DGE and reQTLs are generated by different mechanisms, and focusing on the overlap is an unnecessarily narrow view.

An unappealing thought is that opposite sign effects are enriched in false positives, especially as they seem to show no positional enrichment near the TSS. While it is known that stimulation-specific reQTLs are more distal than baseline eQTLs [79], the HIRD reQTLs are evenly spread across the *cis* window. Some reQTLs may be statistical artifacts of the shrinkage of effects by `mashr`. Small and opposite effects generated by noise may be frequent enough for `mashr` to consider them a “pattern” of effects. This might explain the clear separation of the distribution of *z*-statistics for difference in beta between reQTLs and shared eQTLs. Conversely, it may be that small and opposite effects are more prevalent than expected, and combining `mashr` and a difference in betas test is the best framework for detecting them. To confirm either way, it may be necessary to repeat the reQTL calling without the influence of `mashr` shrinkage in a different modelling framework, such as one using a timepoint-genotype interaction term [96]. A complementary approach for validating these opposite sign reQTLs using the existing RNA-seq data might be within-individual allele-specific expression (ASE) (e.g. RASQUAL [324], PLASMA [325]). One would expect a true opposite sign reQTL effect to be recapitulated as opposite directions of allelic expression imbalance between timepoints. ASE may also provide more interpretable effect sizes than eQTL betas [326], for purposes such as clustering effect sizes to determine patterns of effects across timepoints [323].

At least one reQTL signal was plausibly not a false positive. The strongest reQTL detected at day 1 was for *ADCY3*, a membrane-bound enzyme that catalyses the conversion of ATP to the second messenger cAMP [327]. *ADCY3* is upregulated after the differentiation of monocytes induced by beta-glucan—into macrophages in a state of trained immunity: a state in which they are more responsive to future immune stimuli [328]. GWASs have implicated the *ADCY3* locus in diseases such as obesity [327] and IBD [180]. *ADCY3* has also been identified as a post-stimulation reQTL in other studies involving stimulated blood immune cells: in PBMCs 24 h after *in vitro* infection with rhinovirus [83], *in vivo* in whole blood at day 1 after vaccination with seasonal TIV [94], and in whole blood after *in vitro* stimulation with *M. leprae* antigen for 26–32 h [92]. Given the diversity of stimulations and tissue types, the effect is likely a consequence of general immune activation, rather than a Pandemrix-specific response.

The strength of the *ADCY3* reQTL at day 1 was found to be modified by `xCell` estimates of monocyte abundance. The `xCell` scores are imperfect. Compared to FACS measurements in a cohort subset, the `xCell` scores were only weakly correlated, and the signatures used by `xCell` may be less accurate after vaccine stimulation. Fortunately, statistical colocalisation confirmed that the day 1 *ADCY3* reQTL signal is likely to be a monocyte-specific effect—and independent to the IBD signal in the locus, which colocalises with CD4⁺ T cell eQTL datasets. The proportion of monocytes in the PBMC compartment increases at day 1, supported by both FACS [162] measurements and an increase in monocyte `xCell` score. Expression of *ADCY3* in HIRD is not

monocyte-specific: despite the increase in monocyte proportion, no upregulation was observed at day 1. Colocalisation was also not restricted to stimulated monocytes. The probable mechanism is an increased proportion of the bulk sample taken up by monocytes at day 1 providing more monocyte-derived *ADCY3* transcripts, rather than an upregulation-driven increase in detection power, or a vaccine-induced activation of the locus at day 1. Although multi-trait colocalisation proved to be the crucial piece of evidence suggesting the effect is not related to T cells, only 15 immune cell types were included in the analysis, so it is possible the reQTL is not entirely monocyte-specific.

Overall, cell type interactions were only detected at 16/1154 reQTLs. Although power to detect significant interactions is lower than power to detect main effects—not helped by the unclear reliability of xCell scores—it is still likely that mechanisms other than shifts in cell abundance underlie a large number of the detected reQTLs. One type of mechanism by which *cis*-eQTLs affect expression is through their impact on transcription factor (TF) binding affinity to motifs in promoters and enhancers [329]. Immune cells, including monocytes, are heavily regulated by cell type-specific TFs [330]. Cell type-specific expression of TFs has been proposed as a model for explaining magnifying, dampening, and opposite sign reQTL effects; for example, opposite sign effects could result from different TFs regulating the same gene via the same regulatory element, with activating effects in one cell type and suppressive effects in another [76]. There is evidence that TF activity is important for *in vivo* immune reQTLs: Çalışkan *et al.* [83] found rhinovirus reQTLs in PBMCs were enriched in ENCODE chromatin immunoprecipitation sequencing (ChIP-seq) peaks for the TFs *STAT1* and *STAT2*, and Davenport *et al.* [96] found interferon and anti-IL6 drug reQTLs likely disrupt *ISRE* and *IRF4* binding motifs. Rather than condition-specific expression of the eGene, reQTL effects could be generated by condition-specific expression of TFs whose activity is affected by the reQTL variant. A genomic feature enrichment for TF binding sites and other regulatory elements among (fine-mapped) HIRD reQTL variants could expose shared regulatory factors that explain subsets of the remaining reQTLs. This would also help evaluate if the even distribution of reQTLs across the *cis* window is a cause for concern.

Not only are the mechanisms at many detected reQTLs unknown, there may be many more reQTLs yet to detect in HIRD. Multiple independent eQTLs are present for a large fraction of eGenes [331]. As a single lead variant for reQTL assessment was chosen per gene to avoid reQTLs caused by differential tagging efficiency, I could not detect secondary reQTLs masked by a stronger shared eQTL for the same gene. This is not expected to be uncommon, as the effective sample size for shared eQTLs is usually large due to borrowing of information across conditions. Secondary eQTL signals tend to be weaker, more distal to the TSS, more likely to be enriched in enhancers rather than promoters, and—importantly—more context-specific [55, 85, 332, 333]. The proportion of genes with reQTLs I detect based only on the lead signal likely represents a lower bound. Stepwise conditional analyses at each lead variant will be required to uncover secondary associations, which then can be compared across timepoints in the same manner as the primary associations. These associations, although weaker on average, may actually have more variable effects between timepoints. I also did not consider *trans*-eQTLs due to sample size, which are more likely to be condition specific than *cis*-eQTLs [56, 59, 79].

Finally, I address the prospect that common genetic variation may explain some variation

in antibody response to Pandemrix. I have indirectly demonstrated genotype-dependent effects on expression response by identifying reQTLs with differing effect size between timepoints, but have yet to determine resulting genotype-dependent differences in antibody phenotypes. Some of the identified reQTLs will undoubtedly affect genes whose expression or post-vaccination expression change correlates with antibody response, but correlation is not transitive [334], so correlation of genotype with expression and expression with antibody response does not imply a correlation between genotype and antibody response. Formal tests such as the CIT [263] will be required to distinguish mediation of genotype-antibody response associations through gene expression from competing models. Franco *et al.* [94] realised this, but concluded that they had insufficient power for the CIT, with a greater sample size and comparable study design to HIRD. The HIRD cohort is also too small for a direct GWAS of Pandemrix antibody response. An approach for prioritising reQTLs that contribute to the antibody response to Pandemrix may need to leverage external genetic associations with similar phenotypes found in larger cohorts; for example, colocalisation with existing GWAS summary statistics for antibody response to other vaccines (ideally adjuvanted and inactivated vaccines). However, due to the number of possible generating mechanisms for bulk reQTLs *in vivo*, careful interpretation will be required to glean any insight into the biology of the stimulation in question. Chapter 5 will continue the discussion on the methodologies, experimental designs, and upcoming technologies required to complement the *in vivo* reQTL study design.

Chapter 4

Transcriptomic associations with anti-TNF drug response in Crohn's disease patients

The work presented in this chapter is a collaboration between the Wellcome Sanger Institute, the Royal Devon and Exeter Hospital National Health Service (NHS) Foundation Trust, the University of Exeter, and AbbVie. I would like to thank Nicholas Kennedy, Tariq Ahmad, and the AbbVie team for kindly extending the opportunity to collaborate on the PANTS cohort; Mark Reppell, Samantha Lent, and others at AbbVie, for performing the RNA-seq library preparation and sequencing, initial quality control, alignment and quantification, and estimation of cell proportions from methylation data; Simeng Lin, for advice on the sample structure of PANTS; Aleksejs Sazonovs, for performing the genotype quality control; and other individuals in the Sanger-AbbVie-Exeter PANTS working group, for their feedback during our video conferences.

4.1 Introduction

4.1.1 Crohn's disease and inflammatory bowel disease

Crohn's disease (CD) is a chronic inflammatory disease of the gastrointestinal tract. Along with ulcerative colitis (UC), it is one of the two main forms of inflammatory bowel disease (IBD). CD is characterised by patchy inflammation, where lesions are interspersed with regions of normal mucosa. The lesions can be distributed anywhere in the gastrointestinal tract, and tend to be transmural, affecting all layers of the gut wall. In contrast, UC is characterised by continuous inflammation, with lesions that are superficial rather than transmural, and restricted to the colon [335]. Whilst the two are distinct forms of IBD, similarities in clinical presentation, available therapies, and genetic architecture mean they have often been studied together. Both are immune-mediated inflammatory diseases (IMIDs), a group of related diseases involving immune dysregulation of common inflammatory pathways. Other IMIDs include type 1 diabetes (T1D), systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), multiple sclerosis (MS), and psoriasis [336, 337].

Pathogenesis of CD is not completely understood, but involves interaction of the immune

system, environmental factors (e.g. smoking, stress, diet [335, 338]), and gut microbial factors in a genetically-susceptible individual [339]. Since the seminal discovery by linkage analysis in 2001 that genetic variation in *NOD2* is linked to **CD** risk [340], much progress has been made in establishing the disease's genetic architecture. The most recent **genome-wide association study (GWAS)** studies catalogue over 240 risk loci for **IBD** [180]. Most associations are shared between **CD** and **UC**, but there is strong heterogeneity of effects at some loci, such as *NOD2* being only associated with **CD** risk [341, 342].

CD has historically been considered a disease of the Western world. The highest prevalence and incidence of new **CD** cases are in North America and Western Europe [335], although disease burden is now rising in newly industrialised countries in Asia, Africa, and South America [343, 344]. The modal age of onset is typically between late adolescence and early adulthood. The disease is progressive: within 10 years of diagnosis, approximately 50 % of **CD** patients develop further complications (strictures or penetrating lesions); within 20 years, approximately 15 % will require surgical intervention [335]. Given the rising prevalence and large impact on quality of life, there is active research into developing treatment regimens with the goal of inducing complete mucosal healing [335, 345].

4.1.2 Anti-TNF therapies for Crohn's disease

Tumour necrosis factor (**TNF**), also known by the archaic name **TNF- α** , is a proinflammatory cytokine produced mainly by immune cells such as monocytes, macrophages, natural killer (NK) cells, T cells, and B cells. It is synthesised in transmembrane form, then enzymatically cleaved into its soluble form. **TNF** binds to receptors TNFR1 and TNFR2; most cells in the body express one receptor or the other. Binding triggers a signalling cascade that in different contexts regulates inflammation, apoptosis, cell proliferation, and cell survival [346–348]. In the context of **IBD** pathogenesis, current models suggest high **TNF** levels promote apoptosis of monocytes, macrophages, and gut epithelial cells via TNFR1, while inhibiting apoptosis of mucosal CD4 $^{+}$ T cells via TNFR2 [345, 348, 349], overall encouraging maintained gut inflammation.

The development of anti-**TNF** biologic therapies has revolutionised patient care for **CD** and a number of other **IMIDs** in the last two decades. Infliximab and adalimumab are the two major anti-**TNF** drugs in use. Both are IgG1 monoclonal antibodies that bind both soluble and transmembrane **TNF**, inhibiting their interactions with **TNF** receptors [349, 350]. Two main mechanisms for their action have been proposed: induction of CD4 $^{+}$ T cell apoptosis in the gut mucosa by inhibiting the **TNF**-TNFR2 interaction; and binding of the antibody tail (Fc) region of the drug to Fc receptors on monocytes, inducing their differentiation into wound-healing M2 macrophages [345].

Adalimumab is a human antibody, typically administered subcutaneously via auto-injector pen, with two initial doses aimed to induce remission, then a dose every two weeks to maintain remission. Infliximab is a chimeric mouse-human antibody, administered via intravenous infusion, with a three-dose induction, then doses every eight weeks for maintenance [349]. Anti-**TNF** biologics consistently rank among the drugs generating the highest global revenues. In 2017, spending on adalimumab (Humira) in developed markets was estimated at 20.7 billion USD—almost double the spending on second-ranked insulin glargine (Lantus, 10.5 billion USD) [351].

4.1.3 Anti-TNF treatment failure

Unfortunately, anti-TNF therapy is not always effective at treating CD. Various types of treatment failure can occur: primary non-response (PNR) within the induction period (the first 12–14 weeks for adalimumab and infliximab), developing secondary loss of response (LOR) during maintenance after an initial response, failure to achieve remission after the treatment course, and adverse events that lead to treatment stoppage [352]. For IBD patients, the incidence of PNR is 10–40 %, and the incidence of secondary LOR is 24–46 % in the first year of treatment [353–355]. Another factor affecting treatment outcome is immunogenicity, the generation of antibodies against the drug, thought to increase the probability of treatment failure and LOR by increasing drug clearance rate [350, 355]. As a chimeric antibody, infliximab is more immunogenic than adalimumab [355, 356]. Although remission with complete mucosal healing remains the gold standard for treatment success or failure [335], PNR and LOR phenotypes can be defined much earlier in the treatment course, and help guide changes in treatment regimens, such as dose intensification or switching to a drug class with a different mechanism of action [350, 353].

Anti-TNF biologics are near the top of the therapeutic pyramid for CD in the UK, among the treatment options with the highest toxicity and costs [357]. The traditional approach to disease management in the UK is “step-up”, beginning at the bottom of the pyramid with steroids [349, 357]. This may undertreat patients that require more aggressive therapies, allowing the disease time to progress. An inverted approach begins at biologic therapies, then steps down the pyramid if possible. This risks exposing patients to aggressive therapies they may not have needed [354]. The best approach would be to predict whether a particular treatment will be required and effective for a patient, especially given the costliness and patient risks associated with therapies near the top of the pyramid. Reliable baseline prediction would be especially valuable for stratifying patients to specific therapies from treatment initiation.

4.1.4 Predicting patient response to anti-TNFs

Clinical variables reported to have associations with anti-TNF response include age, disease duration, body mass index (BMI), smoking, C-reactive protein (CRP) levels, faecal calprotectin levels, serum drug concentrations, and anti-drug antibody concentrations. These associations have mostly been found in small retrospective cohorts, and have rarely been independently validated [348, 354, 358–361]. In the Personalised Anti-TNF Therapy in Crohn’s Disease (PANTS) study, the largest study of infliximab and adalimumab response in CD patients to date (enrollment $n = 1610$), baseline obesity, smoking, and greater disease activity were associated with low serum drug concentration after induction. Low drug concentration was in turn associated with PNR and non-remission, suggesting immunogenicity may be mediating treatment failure [355].

Multiple studies have attempted to define transcriptomic predictors for anti-TNF response in gut biopsies and blood [348, 361]. In gut biopsies, expression of sets of “signature” genes were found to be predictive of mucosal healing after infliximab treatment in cohorts of UC (*TNFRSF11B*, *STC1*, *PTGS2*, *IL13RA2*, *IL11*; $n = 46$ [362]) and CD patients (*TNFAIP6*, *S100A8*, *IL11*, *GOS2*, *S100A9*; $n = 19$ [363]). Expression of *OSM* was associated with anti-TNF response defined by improved Mayo score, a multiparameter clinical score of UC activity ($n = 227$)

[364]. Most recently, single-cell RNA sequencing (RNA-seq) identified a module of IgG plasma cells, inflammatory mononuclear phagocytes, activated T cells, and stromal cells associated with clinical remission after anti-TNF therapy in two separate CD cohorts (total $n = 340$) [365].

As obtaining blood samples is non-invasive, there has been great interest in finding transcriptomic predictors of response in blood. While blood is not the main disease-relevant tissue for CD, many genes in gut biopsy signatures have high expression in infiltrating immune cells, and blood gene expression may capture the precursors of those cells [366]. Blood *TREM1* expression has been identified as a marker of anti-TNF response in two studies with inconsistent directions of effect. Gaujoux *et al.* [366] defined response based on “clinical and/or endoscopic improvement”. *TREM1* expression was lower in infliximab responders in gut biopsies (total $n = 72$), but higher in responders in a separate cohort measuring baseline whole blood expression ($n = 22$). Verstockt *et al.* [367] defined response based on endoscopic remission, reporting *TREM1* to be a marker of response with lower expression in responders to infliximab and adalimumab in both baseline gut biopsies ($n = 44$) and baseline whole blood ($n = 54$). Proposed reasons for the discrepancy include false positives due to small sample sizes, differences in patient ethnicity, and different definitions of response [348, 368].

Attempts have also been made to find genetic markers for response. Anti-TNF response does not necessarily share the same genetic architecture as disease risk. Variants in TNF-regulated genes that are also associated with IBD risk (*NOD2*, *TNFR1*, *TNFR2*) are not associated with response to infliximab [348, 361]. A number of candidate gene studies found single nucleotide polymorphism (SNP) associations with response in genes such as apoptosis-related Fas ligand and caspase-9 that have yet to be validated [354, 369]. Recently, larger cohorts have enabled GWASs of anti-TNF response in IBD. In PANTS, although no associations to PNR were genome-wide significant, HLA-DQA1*05 carriage was found to be associated with higher anti-drug antibody levels, which was in turn associated with LOR [370], but larger samples may be needed to find direct associations between HLA-DQA1*05 carriage and LOR.

Overall, small sample sizes and variation among studies in analysis methods, anti-TNF drug, response definition, tissues sampled, and disease make a consensus hard to establish. Few markers of any type—clinical, transcriptomic (gut/blood), or genetic—have been validated in independent studies. No algorithms using such markers for predicting IBD patient response to anti-TNF therapy have yet been translated to clinical practice, although several are currently undergoing validation [361].

4.1.5 Chapter summary

This chapter focuses on identifying novel transcriptomic associations with anti-TNF primary response in a subset of the PANTS cohort with longitudinal RNA-seq data from the first year of follow-up. I model differential gene expression (DGE) between primary responders and non-responders at the gene and module-level at baseline (week 0), post-induction (week 14), and during maintenance (week 30 and week 54). As this is one of the largest datasets currently available for assessing transcriptomic associations with anti-TNF response in IBD, I attempt to validate and resolve conflicts in the literature for previously identified transcriptomic markers such as *TREM1*. Finally, I integrate existing genotype data to map response expression quantitative

trait loci (reQTLs) between timepoints, with the aim of identifying common genetic variants controlling expression response to anti-TNF drugs.

4.2 Methods

4.2.1 The Personalised Anti-TNF Therapy in Crohn's Disease (PANTS) cohort

PANTS is a UK-wide, prospective, observational cohort study of response to anti-TNF therapy in CD patients, described in detail by Kennedy *et al.* [355]. The study was registered with ClinicalTrials.gov identifier NCT03088449, and the protocol is available at <https://www.ibdresearch.co.uk/pants/>. Total enrollment was 1610 patients, who were at least 6 years old, had active luminal CD, and were naive to anti-TNF therapy. Patients were invited to attend up to ten major study visits over a maximum follow-up period of three years, or until drug withdrawal.

The anti-TNF drugs evaluated were adalimumab (ADA) and infliximab (IFX). The study also evaluated infliximab biosimilars; data from patients who received a biosimilar are not included in this chapter. All major visits were scheduled immediately prior to a drug dose. Adalimumab and infliximab have 2-week and 8-week dosing intervals respectively, so the timing of major visits was chosen such that the same visit structure could be used for patients on either drug. Additional visits were scheduled in case of secondary LOR or premature study exit due to drug withdrawal, usually replacing the next scheduled major visit.

The overall rate of primary non-response by week 14 was 21.9 % for infliximab and 26.8 % for adalimumab. The rate of secondary LOR by week 54 among primary responders was 36.9 % for infliximab and 34.1 % for adalimumab. Rate of remission by week 54 was 39.1 % for infliximab and 33.1 % for adalimumab.

4.2.2 Definition of timepoints

The RNA-seq data for this chapter comes from a subset of the cohort sampled around four timepoints: week 0, week 14, week 30, and week 54. These are the target timings for four major visits in the first year of follow-up. The week 0 major visit is the visit immediately prior to the first dose of drug. Week 0 to week 14 is the induction period. After week 14, patients continued to take their drug according to the prescribed schedule (a dose every 8 weeks for infliximab, and every 2 weeks for adalimumab). Whole blood samples at major visits were taken prior to the scheduled drug doses that aligned with those visits, labelled with the visit's name, and preserved for RNA-seq in Tempus Blood RNA Tubes. As sampling was always done prior to a scheduled dose, the measured transcriptome reflects the state at trough drug levels.

I mapped samples from major and additional visits to four discrete timepoints centered around the four major visits. As it could not be guaranteed that visits occurred on the exact day specified in the protocol, I considered the visit windows defined by Kennedy *et al.* [355]: week 0 (week -4–0)*, week 14 (week 10–20), week 30 (week 22–38), and week 54 (week 42–66). Samples

*Samples at negative weeks or study days indicate the patient was first sampled before the day they took the first drug dose (week or day 0).

were mapped to timepoints based on their sample label and study day according to the following mapping criteria:

- Labelled major visit samples were mapped to the corresponding timepoint, regardless of whether they fell within the corresponding window i.e. a sample *labelled* week 54 was always mapped to the week 54 timepoint.
- Samples taken at additional (LOR or exit) visits falling within one of the windows were mapped to that timepoint, unless the patient also had a major visit sample inside that window. This avoided any patient having multiple samples for a single timepoint.

Only a small minority of major visit samples fell outside their corresponding windows, mostly for later timepoints where there was more variation around the target day. Inclusion of samples from additional visits was important, as they often replaced major visits for patients with PNR or LOR. For example, a patient who developed PNR by week 14 and decided to exit the study would not have a labelled week 14 major visit, but may still have a sample taken at that time labelled as an additional exit visit. The mapping of samples to timepoints is shown in Fig. 4.1. Samples included under both of the above mapping criteria should be representative of trough drug levels, as major visits and LOR visits were always scheduled prior to a drug dose, and exit visits were scheduled for when the next drug dose would have been.

4.2.3 Definition of primary response and primary non-response

The definition of primary response and non-response was based on the clinical decision tree from Kennedy *et al.* [355]. Primary response and non-response could be assessed from week 12, with a final classification made by the scheduled week 14 visit. The criteria for primary non-response was *either* of the following:

- exit for treatment failure before week 14 (e.g. as decided by physician global assessment), *or* corticosteroid use at week 14 (a continuing or new prescription);
- compared to week 0, a decrease in CRP by <50% or to >3 mg l⁻¹, *and* a decrease in Harvey Bradshaw index (HBI) by <3 points or to >4.

The criteria used to define primary response was *all* of the following:

- not classified as a primary non-responder;
- compared to week 0, a decrease in CRP by ≥50% or to ≤3 mg l⁻¹, *and* a decrease in HBI by ≥3 points or to ≤4.

Grey zone patients that only partially met the criteria for either primary non-response or response were excluded in this chapter. There were also additional selection criteria used to choose the subcohort of PANTS patients put forward for RNA-seq. Patients were required to be at least 16 years old, and to have an available baseline serum sample. Within the patients on infliximab, there was propensity score matching between primary non-responders and other patients based on baseline immunomodulator use, baseline steroid use, age, sex, and BMI. As PANTS was

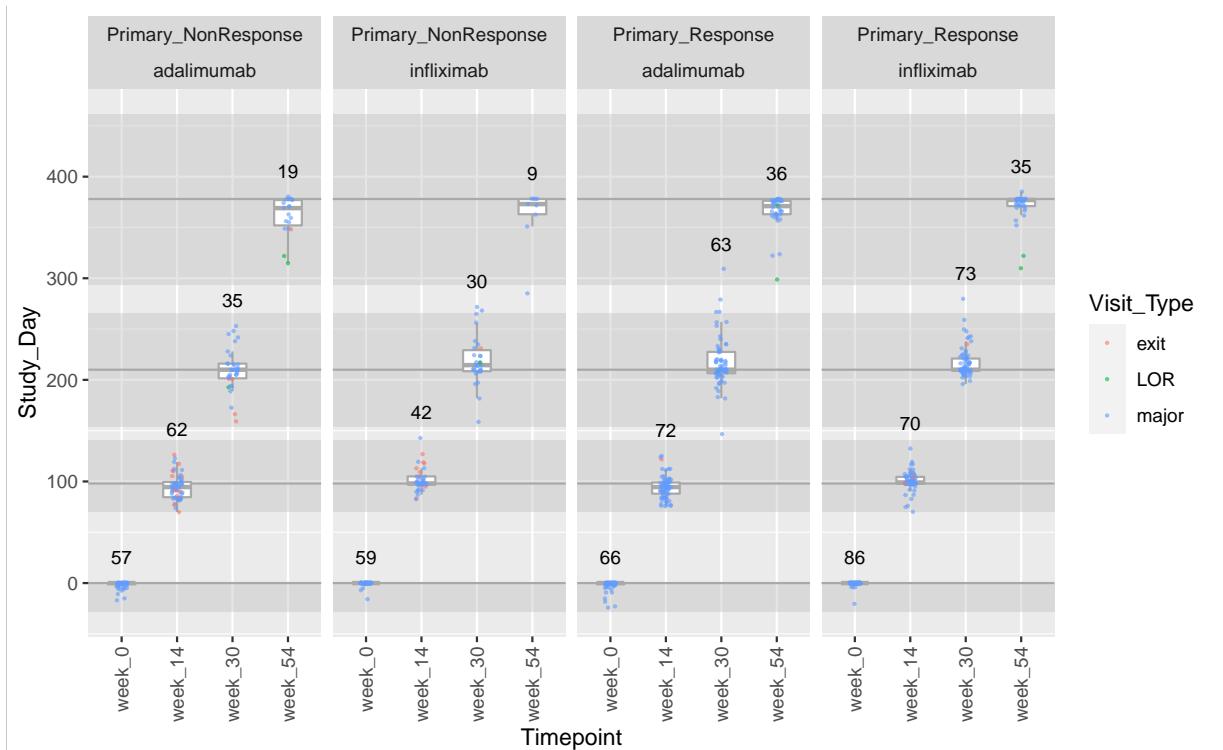


Figure 4.1: Sample size and study day distribution for PANTS study RNA-seq samples, stratified by timepoint and study group. Windows from Kennedy *et al.* [355] for the four major PANTS visits are colored in grey. Samples mostly come from major visits, but a small number of LOR and exit visit samples were included according to the criteria in Section 4.2.2.

an observational study that continued until drug withdrawal, a patient’s clinician may have decided to continue anti-TNF therapy even if a patient demonstrated primary non-response, so it was possible for primary non-responders to remain in the study past week 14. Primary non-responders were selected excluding patients known to be in remission at week 54. Primary responders were selected from patients known to be in remission by week 30 or week 54. The primary non-responders and responders in the RNA-seq subcohort thus represent phenotypic extremes of response.

4.2.4 Library preparation and RNA-seq

Total RNA was extracted following the Qiagen QIAsymphony instrument protocol (RNA Isolation PAX RNA CR22332 ID 2915). RNA was quantified with the ThermoFisher QuBit BR RNA (Q10211), and RNA integrity assessed with the Agilent RNA ScreenTape assay (5067-5579, 5067-5577, 5067-5576) on the Agilent 4200 TapeStation.

Library preparation was performed using the Kapa mRNA HyperPrep Kit, including enrichment for messenger RNA (mRNA) using magnetic oligo-dT beads, depletion of ribosomal RNA (rRNA) and globin mRNA using the QIAseq FastSelect RNA Removal Kit, and adapter ligation with IDT xGEN Dual Index UMI adapters. Libraries were sequenced on the Illumina HiSeq 4000 with 75 bp paired-end reads.

4.2.5 RNA-seq quantification and preprocessing

A total of 1141 samples from 396 patients were sequenced to a median depth of ~20 million read pairs. Sequencing data was demultiplexed with Picard*. Sequence quality, overrepresented sequences, adapter content, and sequence duplication rates were checked using FastQC [371]. Reads were mapped to GRCh38 using STAR (v2.6.1d) [372] and deduplicated to unique reads using UMI-tools [373]. Gene expression was quantified against the Ensembl 96 gene annotation with `featureCounts` (v1.6.4) [374].

Samples were filtered to remove outliers (>2 standard deviations from the mean) according to percentage of aligned reads in coding regions reported by Picard, percentage of unique reads, and number of unique reads. Samples that could not be mapped to a timepoint according to Section 4.2.2 were removed. Samples with sex mismatch were removed. Samples from patients with grey zone primary response were removed. Samples for which there was missingness in the data matrix for variables considered in the variable selection process (Section 4.2.6.1) were removed. A total of 814 samples remained after filtering. The number of samples mapping to each timepoint as defined in Section 4.2.2 is shown in Fig. 4.1. The number of samples per patient ranged from one to four, with a median of three (Fig. 4.2).

The Ensembl 96 gene annotation contains 58 884 genes, many of which are not expressed in whole blood. Effective library sizes were computed using the trimmed mean of M-values (TMM) method in `edgeR` [203], then between-sample normalisation for library size was performed using `edgeR:::cpm`, converting counts to counts per million (CPM). Genes with low expression were filtered, requiring >1.25 CPM in $>10\%$ of samples (1.25 CPM being approximately 10 counts at the median library size of 8 million unique mapped read pairs) and non-zero expression in $>90\%$ of samples. Globin genes and short non-coding RNAs (ncRNAs) were removed. A total of 15 511 genes remained after filtering. Finally, CPMs were converted to the \log_2 scale, and precision weights to account for the expression mean-variance relationship were computed for each gene and sample using `variancePartition::voomWithDreamWeights` [375].

4.2.6 Differential gene expression

4.2.6.1 Variable selection by variance components analysis

For each gene, the DGE model was a regression expressing the response variable (gene expression), as a linear function of predictor variables of interest (primary response status, drug, timepoint), and other selected predictor variables. In estimating the association of predictor X to response Y by regression, adjustment for a third variable Z can increase, decrease, or even reverse the effect estimate for X (the regression coefficient). I aimed to select third variables for inclusion into the DGE model that were covariates, defined here as a Z that is associated with Y , but not with X . Such variables are also known as neutral controls [376], precision variables, or prognostic variables. At the cost of one degree of freedom (df), Z explains variation in Y that would otherwise be considered residual, so conditioning on Z increases the efficiency of estimating the effect of X on Y , but does not change the effect estimate.

*<https://broadinstitute.github.io/picard/>

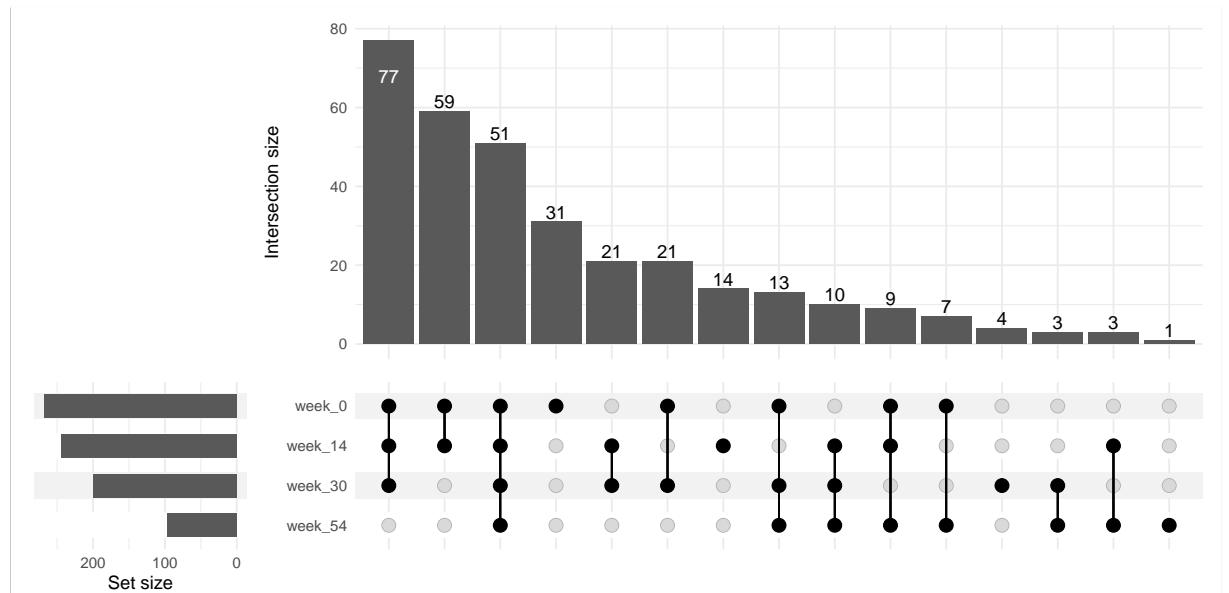


Figure 4.2: Distribution of RNA-seq samples from each patient among timepoints.

Many variables were available for selection; Fig. 4.3 shows their correlation matrix. These included three variables associated with primary response in Kennedy *et al.* [355]: baseline immunomodulator use, smoking, and **BMI**. Also available were proportions of six common cell types in whole blood (CD4⁺ T cells, CD8⁺ T cells, B cells, **NK** cells, monocytes, granulocytes), estimated using the Houseman method (`minfi::estimateCellCounts` [377]) from whole blood Illumina MethylationEPIC methylation array data collected from the same patients and time-points. The Houseman method uses differentially methylated regions between immune cell types as cell type markers [378].

A variance components analysis was performed to quantify the proportion of expression variance explained by each variable for each gene using `variancePartition` [375]. Variables that do not explain much variation in the response are unlikely to improve efficiency if conditioned on. The model was a mixed effects regression model with variables in Fig. 4.3 included as predictors. Additional categorical variables were included for patient and **RNA-seq** library preparation plate. An additional continuous variable consisting of random numbers drawn from the standard normal distribution was included as a null. Granulocyte proportion estimates were dropped to relieve perfect multicollinearity. Categorical variables were coded as random intercepts, and continuous variables as fixed effects. Surprisingly, simulations from Hoffman *et al.* [375] showed variance proportion estimates were unbiased even when coding categorical variables with as few as two categories as random effects, as long as model parameters were estimated using **maximum likelihood (ML)** rather than **restricted maximum likelihood (REML)**. It was also shown this approach avoids overestimates of variance proportions that occur if categorical variables with many levels are treated as fixed.

As downstream **DGE** methods require the same set of predictors for all genes, I aimed to select variables that explained a lot of variance for many genes. Variables that explained the most variance on average were patient, cell proportions, and **RNA-seq** plate (Fig. 4.4). Some variables that did not explain more variance on average than the null nevertheless had high maximum

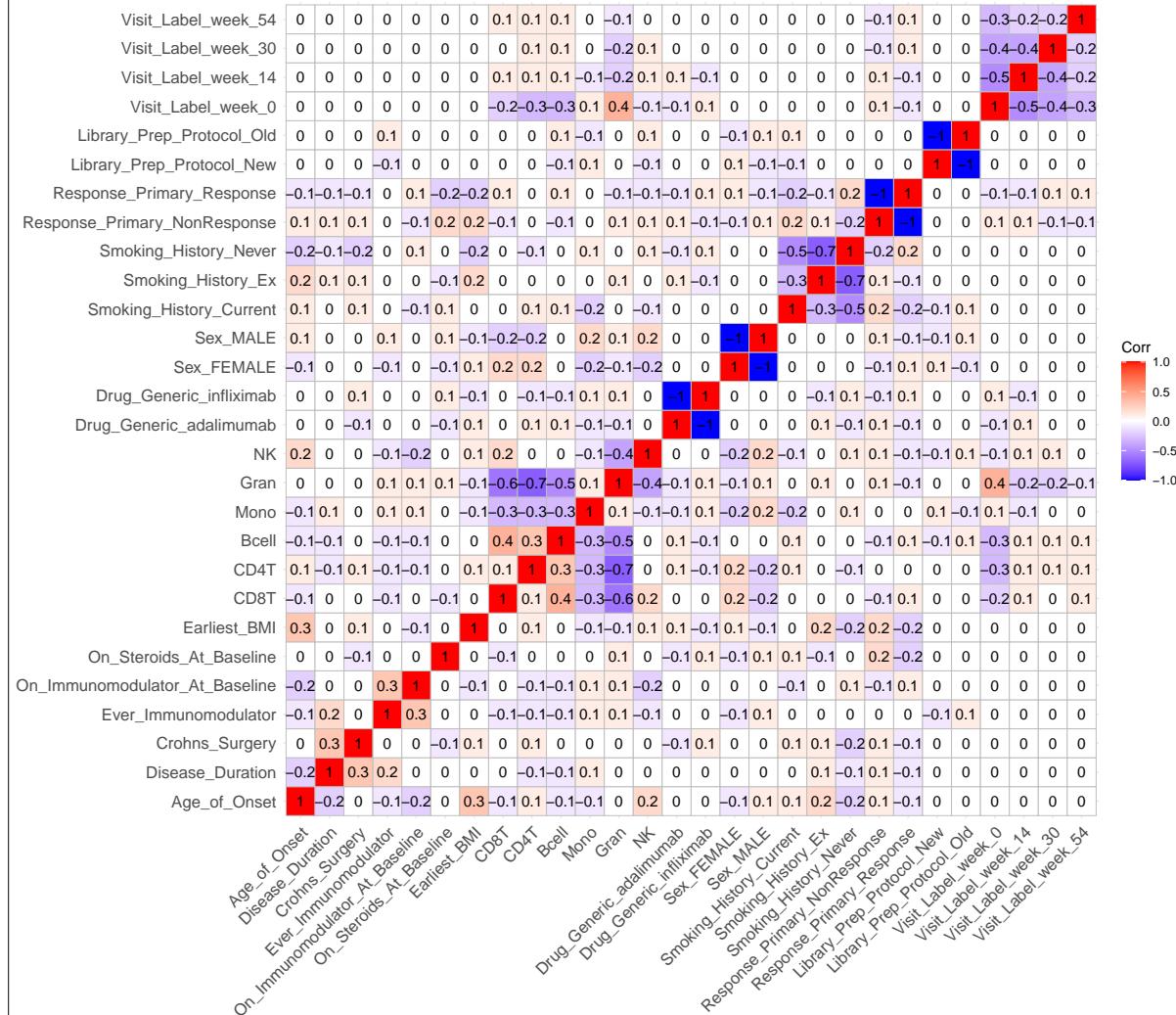


Figure 4.3: Correlation matrix of variables measured in PANTS that were considered as potential predictor variables. NK = NK cell, Gran = granulocyte, Mono = monocyte, Bcell = B cell, CD4T = CD4⁺ T cell, CD8 = CD8⁺ T cell.

values, indicating their importance for a relatively small number of genes. These included sex, library preparation protocol version, and smoking status. However primary response status—a variable of interest—also fell into this group, so it was difficult to justify excluding all variables with lower median variance explained than the null. Consequently, all non-null variables in Fig. 4.4 were selected as predictors in downstream models apart from “Ever_Immunomodulator” (whether the patient had ever had immunomodulator treatment), as that variable had both low median variance explained and was correlated with baseline immunomodulator use. This is a crude approach, but the sample size is large compared to number of **df** lost by including predictors that may not be relevant for some genes.

How might interpretations of effect sizes of interest be affected by including this suite of other variables, all of which can be considered as third variables? If a third variable Z is not a precision variable, but is also associated with X , conditioning on Z changes the effect estimate of X on Y . The regression model is mathematically agnostic to causal relationships between variables, but distinct types of third variable can be distinguished conceptually by assuming the direction of causal relationships [379]. Conditioning on a confounder ($X \leftarrow Z \rightarrow Y$) reduces bias of the effect estimate, conditioning on a collider ($X \rightarrow Z \leftarrow Y$) induces bias, and conditioning on a mediator in the causal pathway ($X \rightarrow Z \rightarrow Y$) changes the effect estimated by removing the indirect effect mediated by Z , usually biasing the effect estimate towards zero*.

From the variance components analysis shown in Fig. 4.4, cell proportions were among the biological factors that explained the most variance on average; they are one of the largest sources of variation in bulk blood expression data, and are a major driver of transcriptional response to immune perturbations [381]. Thus I decided to fit two sets of separate **DGE** models including and excluding cell proportions as predictors, but otherwise identical. Assuming that cell proportions act as a mediator of the drug’s effect on gene expression, these models have complementary interpretations. In models without cell proportions included, differential expression after drug perturbation could represent up or downregulation on a per-cell basis, but could also come from differences in cell proportions induced by the drug. The estimates from models adjusted for cell proportions are more likely to reflect up or downregulation on a per-cell basis. When comparing expression between responders and non-responders, one might also assume cell proportions can mediate the effect of a patient’s response status on expression†. Analogously, estimates of expression differences between responders and non-responders from the two sets of models also have complementary interpretations: total difference, and per-cell differences not due to differences in cell proportions. Throughout this chapter, I interpret the estimates from both sets of models accordingly.

*It is not easy to determine the direction of bias (positive or negative) for any of these cases in general [380].

†The assumption that response status is a stable property of a patient that can be treated as a predictor/independent variable will be discussed in Section 5.2.

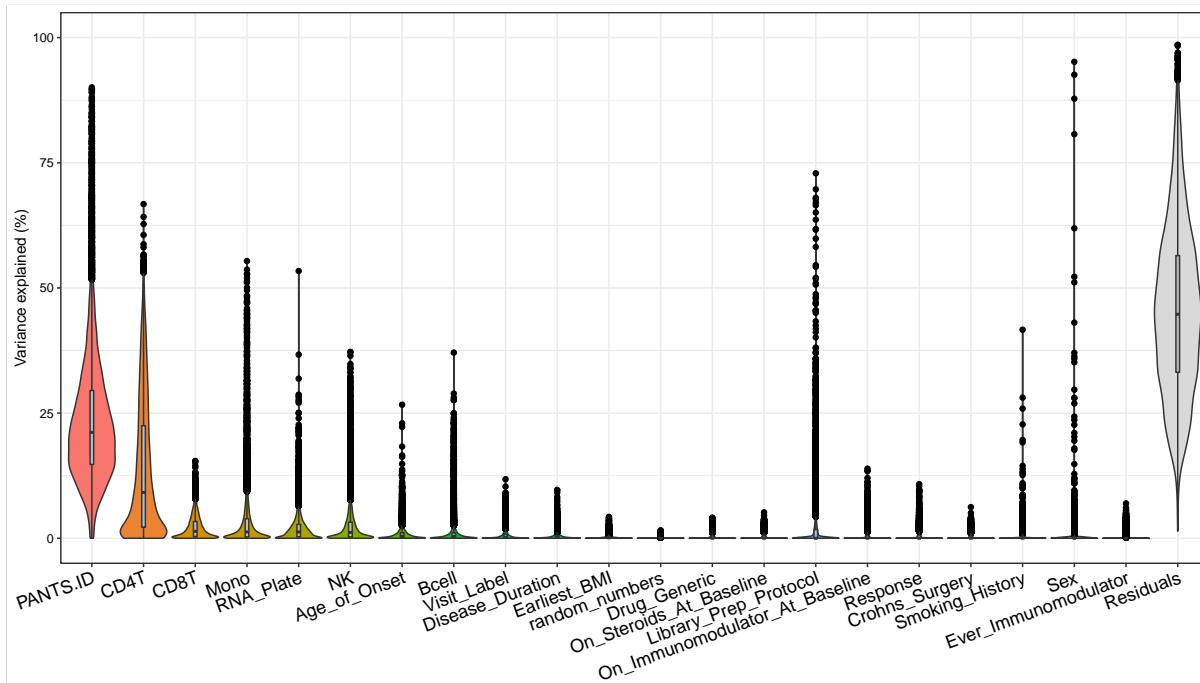


Figure 4.4: Variance components analysis showing the distribution of per-gene percentage of variance in expression explained by each variable. Variables are ordered by the median of per-gene variance explained estimates. random_numbers is a null drawn from the standard normal distribution. PANTS.ID = patient ID, NK = NK cell, Gran = granulocyte, Mono = monocyte, Bcell = B cell, CD4T = CD4⁺ T cell, CD8 = CD8⁺ T cell.

4.2.6.2 Contrasts for pairwise group comparisons

Per-gene DGE models were fit in `dream` [382]. Like the variance components analysis models, these DGE models were linear mixed models:

$$y = 0 + \beta_{trd}G_{trd} + \sum^9 \beta_Z Z + (\sum^5 \beta_C C) + u + v + \epsilon \quad (4.1)$$

where:

- The response variable is gene expression y .
- 0 indicates there is no intercept term.
- G_{trd} is a fixed effect for experimental group defined by combinations of the predictors of interest: timepoint (week 0, 14, 30, 54), response (responder, non-responder), and drug (infliximab, adalimumab). This is equivalent to having an intercept term and a three-way interaction between visit, response, and drug, including all lower order terms, but is more convenient for testing pairwise expression differences between groups, as the coefficient for each term is the estimate of mean expression for that group.
- $\sum^9 \beta_Z Z$ are the non-cell proportion fixed effects chosen in Section 4.2.6.1: sex (Sex), age of disease onset (Age_of_Onset), disease duration (Disease_Duration), smoking history (Smoking_History: ex, current or never), whether the patient has had surgery for CD (Crohns_Surgery), whether the patient was on immunomodulator at baseline

(On_Immunomodulator_At_Baseline), whether the patient was on steroids at baseline (On_Steroids_At_Baseline), **BMI** at baseline (Earliest_BMI), and library preparation protocol version (Library_Prep_Protocol).

- $\sum^5 \beta_C C$ are the cell proportion fixed effects chosen in Section 4.2.6.1, for **NK** cells, monocytes, B cells, CD4 $^+$ T cells, and CD8 $^+$ T cells.
- u is a random intercept for **RNA-seq** plate (**RNA_Plate**).
- v is a random intercept for patient (**PANTS.ID**), nested inside **RNA-seq** plate.

As the interest is in estimating a single coefficient for each predictor's effect size on expression (rather than estimating variance components), most predictors above are modelled as fixed effects. Since **RNA-seq** plate and patient are nuisance variables with a large number of levels, they are modelled as random intercepts. A total of four sets of per-gene models were fit, with and without the cell proportion terms $\sum^5 \beta_C C$, and replacing $\beta_{trd} G_{trd}$ (separate drug models) with $\beta_{tr} G_{tr} + \beta_{dd}$ (pooled drug models) or not. Unlike with variance components analysis, to avoid small-sample bias in estimates of fixed effect standard errors, **REML** was used for estimation [383].

Specific hypotheses were tested using sum-to-zero contrasts, which are linear combinations of model coefficients with weights summing to zero. For example, to test for **DGE** between responders and non-responders to infliximab at baseline in the non-pooled model, I used a contrast where the weight for the week 0/responder/infliximab group coefficient was 1, the weight for the week 0/non-responder/infliximab group coefficient was -1, and all other coefficient weights were 0. To get p -values, the contrast divided by its standard error was compared to the t -distribution using the Satterthwaite approximation for **df**. **False discovery rate (FDR)** was controlled with the **Benjamini-Hochberg (BH)** method, with threshold set at 0.05, computed separately for each contrast*.

4.2.6.3 Spline model of expression over time

The aim was to use expression data from all four timepoints to find genes associated with response, while avoiding a large number of pairwise comparisons. I fit a natural cubic spline (**splines::ns**, R 3.6.2) to the study day to allow for non-linear trajectories of expression over time. A cubic spline is a continuous function defined piecewise in each successive interval between a set of k knots in the range of the input variable. The $k - 1$ pieces between knots are polynomials of degree 3. For a natural spline, the function is constrained to be linear outside of the boundary (first and last) knots to avoid unpredictable behaviour at the boundaries [384]. I set two inner knots at week 14 and week 30, as expression is expected to change after each drug dose. To include all data within the boundaries, the two boundary knots were set at the minimum and maximum values of study day rather than week 0 and week 54. A basis matrix [384] was computed with **ns(Study_Day, knots=7*c(14, 30))**, which is a matrix with 3 columns, each column being a transformation of the input, study day. The columns are fit in the regression model in place

***FDR** could also have been computed globally over all contrasts if it were necessary to have the same t -statistic threshold for statistical significance in all contrasts.

of study day to allow for non-linear effects of study day on expression. The model form used was as in Eq. (4.1), except with $\beta_{trd}G_{trd}$ replaced by $\beta_{rr} + \sum^3 \beta_{bb} + \sum^3 \beta_{rb}rb + \beta_{dd}$, where r is response status, d is drug, $\sum^3 \beta_{bb}$ are the three columns of the basis matrix, and $\sum^3 \beta_{rb}rb$ are the second-order interaction terms between response status and the basis matrix columns. Separate sets of per-gene models were again fit with and without cell proportions $\sum^5 \beta_{CC}$.

When testing for response-associated differences in the spline parameters, the predictors of interest are the interaction terms $\sum^3 \beta_{rb}rb$. The three terms were tested jointly with an F -test, and FDR correction was performed with the BH method, with the threshold set at 0.05. A significant result indicates a significant difference in the trajectory of expression over study day between responders and non-responders.

4.2.6.4 Clustering expression over all timepoints

I clustered genes by their expression trajectories to define sets of genes with similar trajectories over time. This was done to aid the interpretation of significant genes from the cell proportion-adjusted spline model using gene set enrichment analysis. Expression data was converted to the CPM scale using TMM normalisation factors, then regressed against cell proportions. Residuals were centered and scaled per gene. A distance matrix was computed using $1 - r$ as the distance metric, where r is the Pearson correlation. Hierarchical clustering was performed with complete agglomeration for inter-cluster distance (`fastcluster::hclust(method = "complete")`, [385]). The optimal number of clusters was assessed by the gap statistic (`factoextra::fviz_nbclust(method = "gap_stat", nboot = 500)*`), which determines when the change in within-cluster dispersions are no longer significantly improved by increasing the number of clusters [386]. The default `firstSEmax` criteria was used to choose the optimal number of clusters k , which finds the first local maximum at m clusters where $\text{Gap}(m) \geq \text{Gap}(m+1)$, then finds the smallest $k : 1 \leq k \leq m$ such that $\text{Gap}(k)$ is not less than $\text{Gap}(m)$ minus the bootstrapped standard error of $\text{Gap}(m)$. The hierarchical clustering tree was then cut into k clusters.

4.2.6.5 Gene set enrichment analyses

Rank-based gene set enrichment analyses were conducted using `tmod::tmodCERNOtest` [241] and blood transcription modules (BTMs), as described in Section 2.2.10. For each contrast, as the t -statistics are not comparable between genes due to the use of approximate df, I ranked genes by the signed z -score reported by `dream`, which is a monotonic transformation of the p -value. Similarly, moderated F -statistics from the spline model are not comparable between genes, so I used the signed F -statistic reported by `dream` from the transformation of the p -value.

Gene set overrepresentation analyses with the hypergeometric test were conducted with `tmod::tmodHGtest` as detailed in Section 3.2.11.

4.2.7 Genotyping and genotype data preprocessing

Genotype data were subsetted from the post-quality control PANTS cohort genotypes generated by Sazonovs *et al.* [370], where the genotyping and preprocessing pipeline is fully described. In brief,

*<https://rpkgs.datanovia.com/factoextra/index.html>

whole blood samples were collected into EDTA tubes at week 0 and genotyped on the Illumina CoreExome genotyping array. Pre-imputation quality control was performed in accordance with de Lange *et al.* [180]. Imputation was performed using the Sanger Imputation Service with the Haplotype Reference Consortium panel. Post-imputation, samples that were non-European, related (proportion identity-by-descent > 0.1875), or were outliers in genotype missingness or heterozygosity rate were removed; SNPs that were poorly imputed (INFO score < 0.4), deviated from Hardy-Weinberg equilibrium (HWE) ($p < 1 \times 10^{-10}$), had high missingness ($>5\%$), or low minor allele frequency (MAF) ($<1\%$ before subsetting) were removed. 7 503 762 SNPs remained after filtering. Genotypes were converted to dosages of the non-reference allele.

4.2.8 reQTL mapping

The overall strategy and methods used were largely identical those to those used in Chapter 3, laid out in Section 3.2. Differences are described below.

4.2.8.1 Computing genotype principal components

Samples were projected onto principal components (PCs) defined by 1000 Genomes Project samples using SNP weights from akt*, confirming that samples were of European ancestry (Fig. 4.5). Here I chose the first five PCs for use as covariates in expression quantitative trait locus (eQTL) mapping downstream, one more than was chosen in Section 2.2.5 for Human Immune Response Dynamics (HIRD) by the Tracy-Widom test. This should be sufficient to adjust for large-scale population structure, as the PANTS cohort is less ethnically diverse than the HIRD cohort. PCs were centered and scaled before downstream use to improve model convergence.

4.2.8.2 Finding hidden confounders in expression data

Between-sample normalisation and variance stabilisation was applied to the counts matrix (DESeq2::vst [237]), resulting in \log_2 scale expression estimates. Akin to Section 3.2, given known factors (response, drug, five scaled genotype PCs, five cell proportions), PEER [182] was used to infer additional hidden factors that explain variance in the expression matrix for a large fraction of genes. This is similar in principle to the variance components analysis carried out in Section 4.2.6.1, except hidden factors can be unmeasured. To maximise efficiency for *cis*-eQTL mapping, the number of PEER factors retained for each timepoint was selected to maximise the number of genes with at least one significant eQTL (eGenes) detected on chromosome 1 (Fig. 4.6). The selected numbers were 25, 20, 15, and 5 factors; for weeks 0, 14, 30, and 54 respectively.

4.2.8.3 Computing kinship matrices

Akin to Section 3.2.3, leave-one-chromosome-out (LOCO) kinship matrices were computed on typed SNPs for each chromosome using LDADK [274].

*<https://github.com/Illumina/akt>

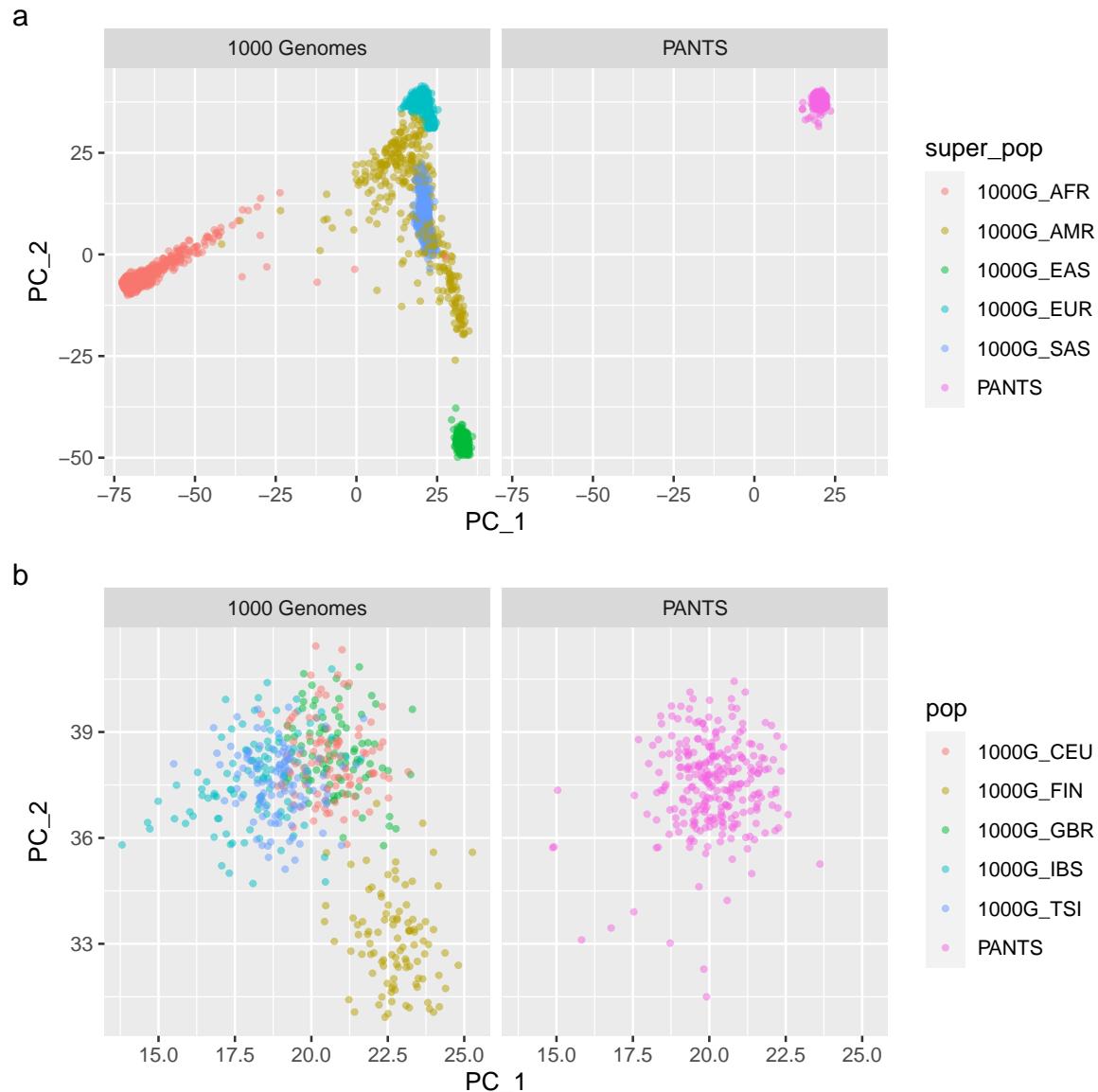


Figure 4.5: 1000 Genomes Project (1000G) samples and PANTS samples projected onto 1000G genotype PC1 and PC2 axes, colored by (a) superpopulation and (b) population. 1000G superpopulations: AFR = African, AMR = Ad Mixed American, EAS = East Asian, EUR = European, SAS = South Asian. 1000G European populations: CEU = Utah Residents (CEPH) with Northern and Western European Ancestry, FIN = Finnish in Finland, GBR = British in England and Scotland, IBS = Iberian Population in Spain, TSI = Toscani in Italia.

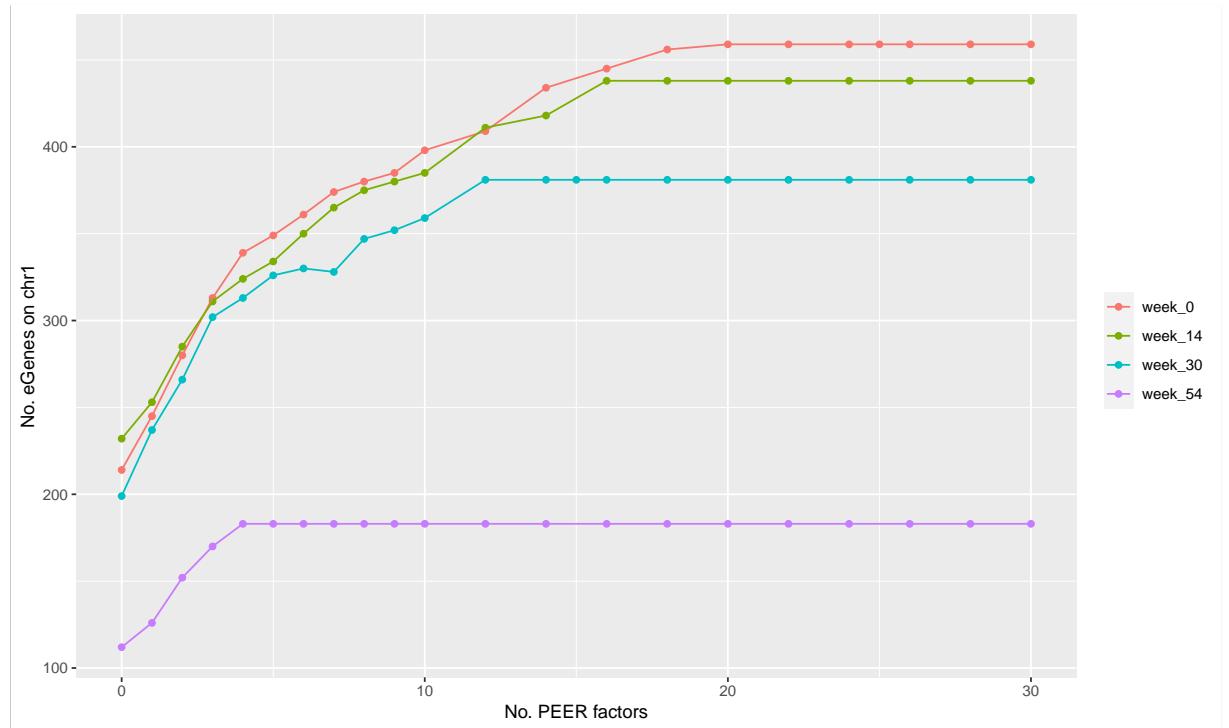


Figure 4.6: Number of eGenes on chromosome 1 vs. number of PEER factors included in eQTL mapping as covariates. FDR computed with hierarchical Bonferroni-BH [305] with significance threshold set at 0.05.

4.2.8.4 Mapping eQTLs per timepoint

As in Section 3.2, eQTLs were mapped in each timepoint using a linear mixed model in LIMIX [273]. The sample sizes with both genotype and expression data available for eQTL mapping at weeks 0, 14, 30 and 54 were 223, 205, 167, and 84 respectively.

For each autosomal gene, *cis*-SNPs within 1 Mbp of the Ensembl gene start (or gene end on the minus strand), were filtered to keep SNPs where the number of samples homozygous for the minor allele was at least five. Small group numbers lead to data points with high leverage that may be unduly influential on the genotype beta. Assuming HWE, this is equivalent to a MAF filter of $\sqrt{5/(223 \times 2)} = 0.11$ in the timepoint with the largest sample size (week 0), and $\sqrt{5/(84 \times 2)} = 0.17$ in the timepoint with the smallest sample size (week 54).

The LIMIX model for each SNP-gene pair had \log_2 expression as the response variable and genotype dosage as the predictor of interest. Other fixed effect predictors were the intercept, known factors (response, drug, five scaled genotype PCs, cell proportions), and PEER hidden factors (timepoint-specific number of PEER factors selected in Section 4.2.8.2). A random intercept term was also included with mean zero and covariance matrix proportional to the LOCO kinship matrix for the SNP's chromosome.

4.2.8.5 Joint reQTL mapping over all timepoints

Analogously to Section 3.2, summary statistics from per-timepoint mapping were input to mashr [281] to map eQTLs jointly over all timepoints. A total of 25 908 527 SNPs were testable at all four timepoints. The null correlation structure of the timepoints was estimated using null

tests within a random subset of 200 000 tests (`mashr::estimate_null_correlation_simple`). Data-driven covariance matrices representing patterns of effects across timepoints were estimated using a strong subset of tests. As the strong subset should contain eQTLs that are likely to have an effect in at least one timepoint, at each timepoint, for each gene with at least one nominally significant eQTL ($p < 0.05$), I selected the eQTL with the smallest p -value, resulting in a strong subset of 129 002 tests. The `mashr` model was then fit on the full random subset in exchangeable Z (EZ) mode, accounting for the computed null correlation and covariance matrices. Finally, posterior betas, standard errors, and local false sign rates (LFSRs) were computed for all tests using the fitted model parameters.

The lead eQTL for each gene was chosen as the eQTL with the lowest LFSR in any condition, breaking ties by highest INFO score, highest MAF, shortest distance to gene start (or end), and smallest genomic coordinate. Each lead eQTL was assessed for being a significant reQTL by a z -test for whether the difference in betas was zero, between the week 0 beta and each of the other three timepoints. Multiple testing for the number of genes was controlled using the BH FDR for each of the three comparisons separately.

4.3 Results

4.3.1 Longitudinal RNA-seq data from the PANTS cohort

To define transcriptomic differences between primary responders and non-responders to anti-TNF therapy in the PANTS cohort, I analysed whole blood RNA-seq gene expression measured at up to four timepoints per patient: week 0 baseline before commencing anti-TNF therapy, and weeks 14, 30 and 54 after commencing anti-TNF therapy. After quality control, expression data was available for 15 584 genes and 814 samples. These samples come from 324 patients, whose characteristics are shown in Table 4.1. The proportion of primary non-responders is high (43.8 %) compared to the overall proportion in the PANTS cohort (23.8 % [355]). This is due to sample selection for RNA-seq to balance the sample size for each combination of drug and primary response status.

4.3.2 Baseline gene expression associated with primary response

Patient primary response to anti-TNF was defined at week 12–14 (after the induction period) according to the clinical decision algorithm from Kennedy *et al.* [355] described in Section 4.2.3, which integrates clinician assessment with changes in CRP level and HBI score. To identify differences in baseline gene expression associated with future primary response, I fit per-gene linear models at 15 511 genes, comparing week 0 gene expression in primary responders with week 0 gene expression in primary non-responders. Comparisons were performed both within infliximab-only and adalimumab-only subgroups, and with both drugs pooled. Models were run both adjusting for cell composition estimates of six immune cell types, and without adjustment. Throughout this section, the significance threshold was set at FDR < 0.05 for each comparison, and positive \log_2 FCs indicate increased expression in responders versus non-responders.

Without adjusting for cell composition, the largest effects were infliximab-only, with 859 genes differentially expressed. Only *KCNN3* (\log_2 FC = −0.84) was significant for the adalimumab-

Table 4.1: Patient characteristics for the PANTS RNA-seq subcohort. Values are count and percentage for categorical variables; mean and standard deviation for continuous variables; *p*-values are for the comparison between drugs.

	adalimumab (ADA)	infliximab (IFX)	drugs pooled	p-value
Sex				0.317
(Col %)				Fisher exact
FEMALE	78 (48.4%)	89 (54.6%)	167 (51.5%)	
MALE	83 (51.6%)	74 (45.4%)	157 (48.5%)	
Age of onset (years)				0.774
Mean (SD)	33.3 (15.4)	32.8 (15.3)	33.1 (15.3)	Wilcoxon rank-sum
Missing	0	0	0	
Disease duration (years)				0.546
Mean (SD)	6.1 (8.1)	5.9 (7.7)	6.0 (7.9)	Wilcoxon rank-sum
Missing	0	0	0	
Smoking status				0.263
(Col %)				Fisher exact
Current	28 (17.4%)	36 (22.1%)	64 (19.8%)	
Ex	55 (34.2%)	43 (26.4%)	98 (30.2%)	
Never	78 (48.4%)	84 (51.5%)	162 (50.0%)	
Crohn's-related surgery				0.549
(Col %)				Fisher exact
FALSE	114 (70.8%)	110 (67.5%)	224 (69.1%)	
TRUE	47 (29.2%)	53 (32.5%)	100 (30.9%)	
On immunomodulator ever				0.543
(Col %)				Fisher exact
FALSE	23 (14.3%)	28 (17.2%)	51 (15.7%)	
TRUE	138 (85.7%)	135 (82.8%)	273 (84.3%)	
On immunomodulator at baseline				0.912
(Col %)				Fisher exact
FALSE	79 (49.1%)	81 (49.7%)	160 (49.4%)	
TRUE	82 (50.9%)	82 (50.3%)	164 (50.6%)	
On corticosteroids at baseline				0.011
(Col %)				Fisher exact
FALSE	113 (70.2%)	92 (56.4%)	205 (63.3%)	
TRUE	48 (29.8%)	71 (43.6%)	119 (36.7%)	
Baseline BMI				0.237
Mean (SD)	25.2 (6.2)	24.3 (5.5)	24.8 (5.9)	Wilcoxon rank-sum
Missing	0	0	0	
Primary response status				0.263
(Col %)				Fisher exact
Primary non-response	76 (47.2%)	66 (40.5%)	142 (43.8%)	
Primary response	85 (52.8%)	97 (59.5%)	182 (56.2%)	
CD8+ T cell (%)				0.380
Mean (SD)	2.8 (4.2)	2.8 (5.2)	2.8 (4.7)	Wilcoxon rank-sum
Missing	38	18	56	
CD4+ T cell (%s)				0.752
Mean (SD)	9.2 (6.3)	9.2 (6.8)	9.2 (6.5)	Wilcoxon rank-sum
Missing	38	18	56	
B cell (%s)				0.094
Mean (SD)	1.9 (2.0)	1.5 (1.9)	1.7 (1.9)	Wilcoxon rank-sum
Missing	38	18	56	
Monocyte (%s)				0.497
Mean (SD)	8.9 (3.5)	9.2 (3.7)	9.0 (3.6)	Wilcoxon rank-sum
Missing	38	18	56	
NK cell (%s)				0.683
Mean (SD)	1.9 (3.2)	1.9 (3.8)	1.9 (3.5)	Wilcoxon rank-sum
Missing	38	18	56	
Granulocyte (%s)				0.911
Mean (SD)	74.3 (9.7)	74.3 (10.8)	74.3 (10.3)	Wilcoxon rank-sum
Missing	38	18	56	

only comparison, and only *SIGLEC10* ($\log_2 \text{FC} = 0.35$) was significant in the pooled analysis (Fig. 4.7). After adjustment for cell composition, there were no longer any significant genes in the infliximab-only analysis, with 856/859 genes that were significant before the comparison having a dampened effect size after correction (smaller absolute effect and same sign), suggesting many effects may be mediated by cell composition. *SIGLEC10* in the combined analysis was also non-significant after adjustment (adjusted $\log_2 \text{FC} = 0.31$, $\text{FDR} = 0.05$). Conversely, at the three genes downregulated in the adalimumab-only analysis that were the only significant genes post-adjustment, I observed increased significance: *PDIA5* (unadjusted $\log_2 \text{FC} = -0.33$, adjusted $\log_2 \text{FC} = -0.35$), *KCNN3* ($-0.84, -0.88$), and *IGKV1-9* ($-1.15, -1.22$).

To identify coordinately up and downregulated gene sets and increase sensitivity for detecting differences between responders and non-responders, I performed rank-based gene set enrichment analyses on the per-gene z -statistics using BTMs: annotated sets of coexpressed genes in peripheral whole blood from Li *et al.* [240] (prefixed “LI”). This module-level analysis was also run both unadjusted (Fig. 4.8) and adjusted for cell composition (Fig. 4.9).

Despite only *SAMD10* having a significantly different effect between drugs at the gene level (a significant interaction between drug and response at week 0), the large global differences observable in Fig. 4.7 were detected in the module-level analysis*. Without adjusting for cell composition, many of the most significantly upregulated modules in the pooled analysis—including upregulation of monocyte (LI.M11.0, LI.S4), neutrophil (LI.M37.1, LI.M11.2), and dendritic cell (LI.M165, LI.S11) modules—appear to be driven by an infliximab-specific effect. These modules had heavily reduced significance after adjusting for cell composition. The new modules that were most upregulated in the pooled analysis after adjustment had more consistent effects between drugs, such as MHC-TLR7-TLR8 cluster (LI.M146), antigen presentation (LI.M71, LI.M95.0), and myeloid cell enriched receptors and transporters (LI.M4.3).

For downregulated modules before adjustment, I observed infliximab-specific effects for NK cell (LI.M7.2) and T cell (LI.M7.0, LI.M7.1) modules. Adalimumab-specific effects were observed for plasma cell, B cell and immunoglobulin modules (LI.M156.0, LI.M156.0, LI.S3), and cell cycle and transcription modules (LI.M4.0, LI.M4.1). After adjustment, the significance of infliximab-specific modules was reduced, but the significance of adalimumab-specific modules and the corresponding interaction effects was increased. For both gene-level and module-level comparisons of baseline expression between responders and non-responders, there is a striking heterogeneity between patients on infliximab and adalimumab that is only partially reduced by cell proportion adjustment.

4.3.3 Assessing previously reported baseline predictors of primary response

In addition to significant genes from this study, Fig. 4.7 is annotated with genes whose expression in gut biopsies or blood has been previously evaluated for baseline prediction of primary response [362, 363, 367, 387]. Some genes expressed in gut mucosa (e.g. *IL13RA2*) were not appreciably expressed in this whole blood dataset, and most other genes that were expressed were

*It is likely the PANTS RNA-seq study is not powered to detect gene-level three-way interaction effects between timepoint, drug and response. I am not aware of which subgroup analyses may have been prespecified during the study design and sample size calculations for the PANTS RNA-seq cohort.

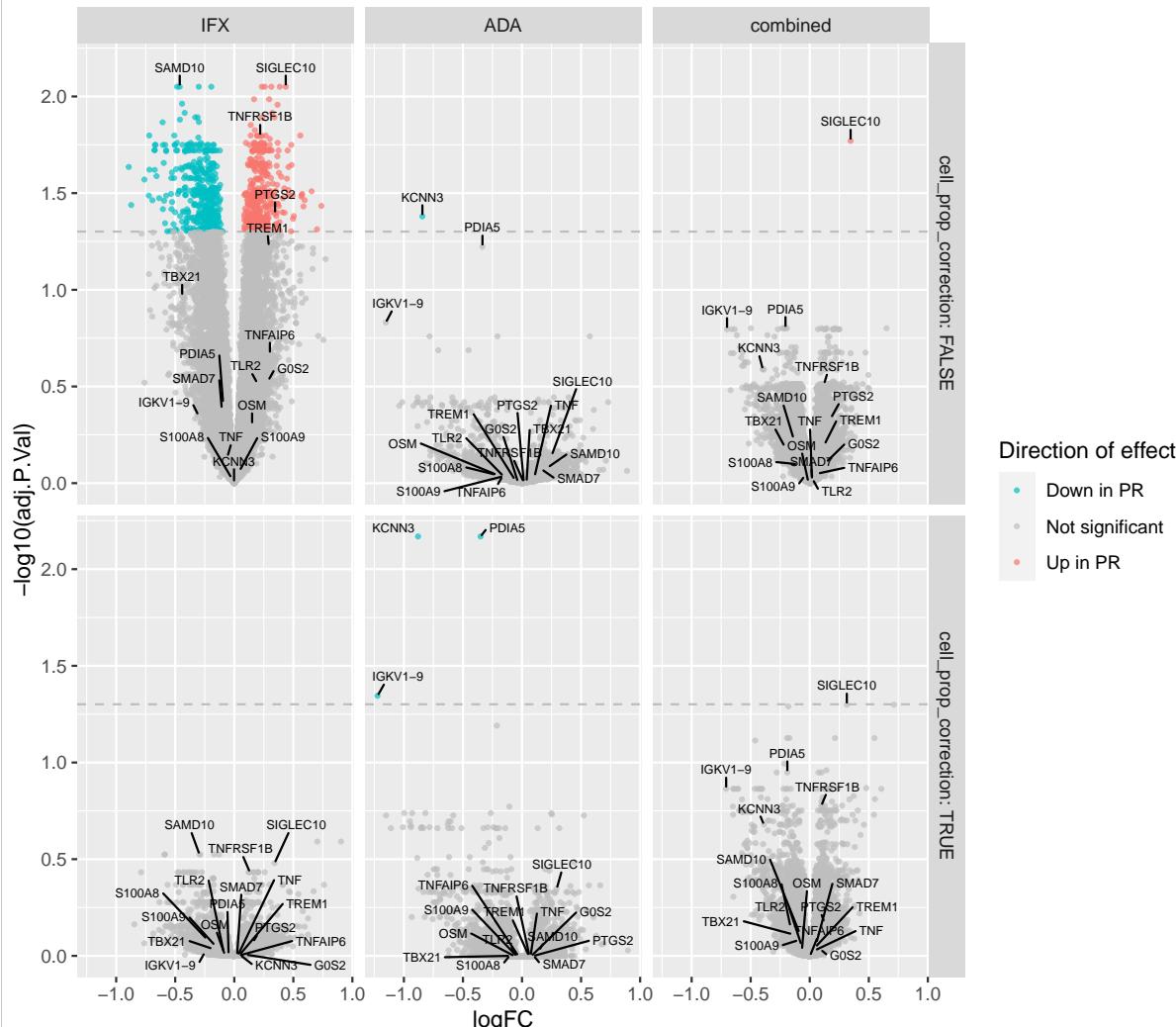


Figure 4.7: Volcano plots of DGE between primary responders (PR) and non-responders at week 0; unadjusted (top row) and adjusted (bottom row) for cell composition; for infliximab (IFX), adalimumab (ADA), or with both drugs pooled. Annotated genes include significant associations from this study and previously reported associations from Section 4.1.4. Dashed line shows significance threshold at $FDR = 0.05$.

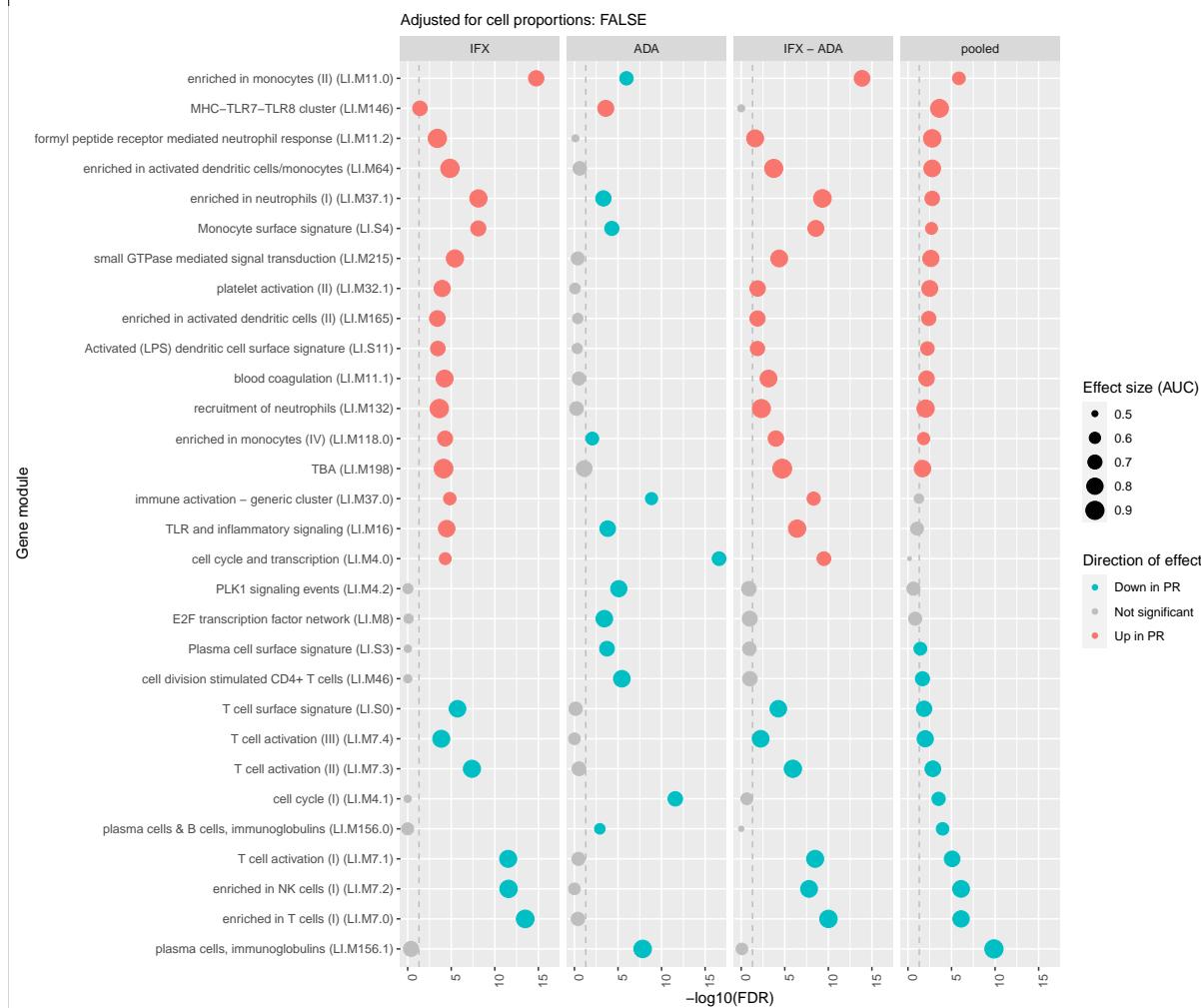
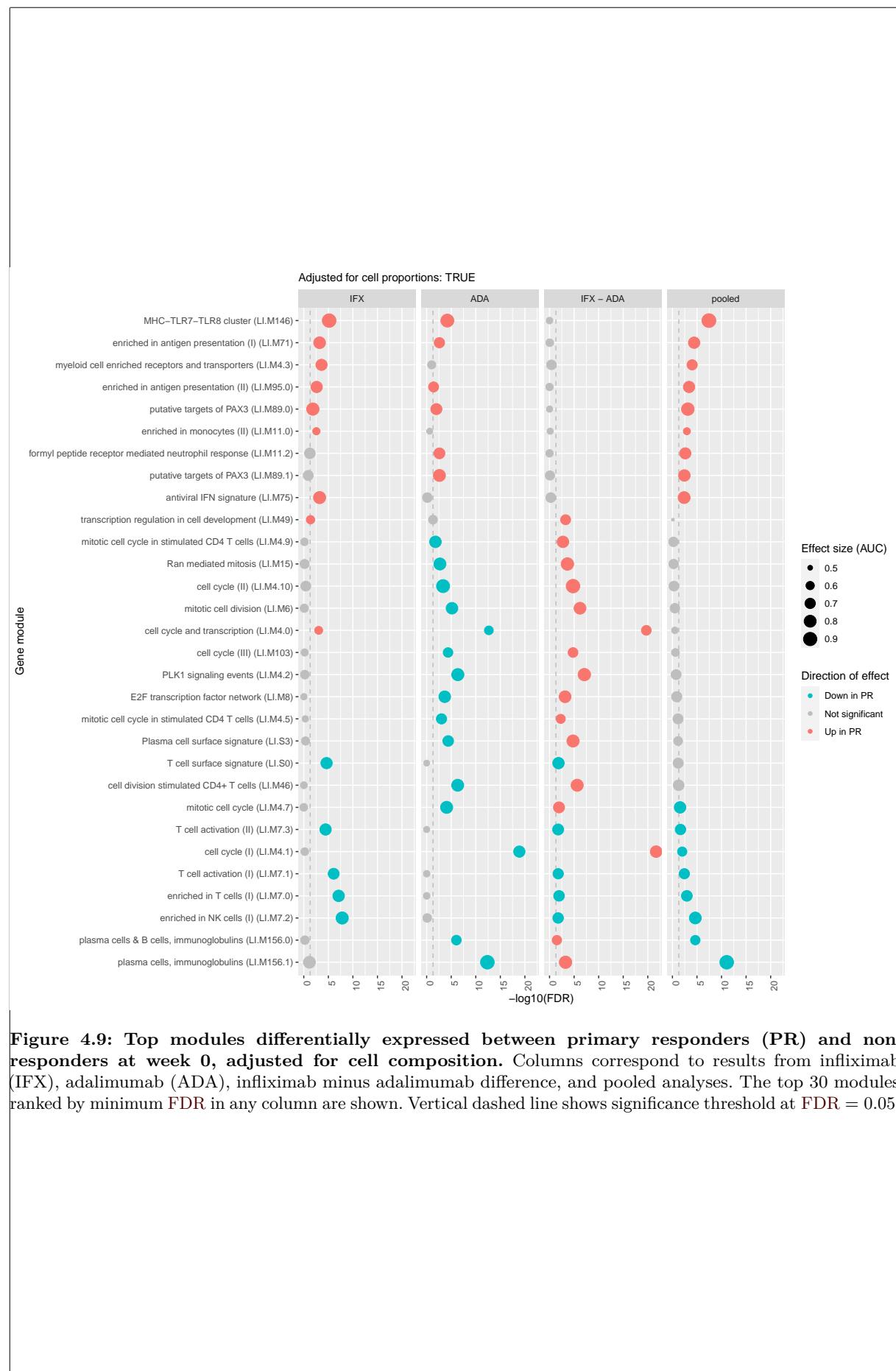


Figure 4.8: Top modules differentially expressed between primary responders (PR) and non-responders at week 0, unadjusted for cell composition. Columns correspond to results from infliximab (IFX), adalimumab (ADA), infliximab minus adalimumab difference, and pooled analyses. The top 30 modules ranked by minimum FDR in any column are shown. Vertical dashed line shows significance threshold at FDR = 0.05.



not significantly differentially expressed. Only *TNFRSF1B* and *PTGS2* were associated with primary response, being upregulated at baseline in responders, specifically in the infliximab-only comparison, unadjusted for cell composition. *TNFRSF1B* was found by Verstockt *et al.* [367] to be downregulated in baseline inflamed mucosal biopsies of responders to anti-TNF therapy in IBD patients ($n = 44$, FC = 0.72, $p = 0.008$). *PTGS2* was found by Arijs *et al.* [362] to be downregulated in baseline mucosal biopsies of responders to infliximab for Crohn's colitis ($n = 46$). The directions of effect in both cases are opposite to this study, but comparisons are hard to draw between blood and mucosal biopsies.

A previously identified marker in blood, *TREM1* was found to have opposing effects in two studies by Gaujoux *et al.* [366] ($n = 22$) and Verstockt *et al.* [367] ($n = 54$). Here, *TREM1* showed the strongest differences between responders and non-responders at baseline in the infliximab subcohort, but did not reach significance before ($\log_2 \text{FC} = 0.29$, FDR = 0.06) nor after adjusting for cell composition ($\log_2 \text{FC} = 0.05$, FDR = 0.99). The sample size in this study for the infliximab subcohort at baseline is $n = 145$ (Fig. 4.1), so it is expected the power in this study is greater.

4.3.4 Post-induction gene expression associated with primary response

The same methodology applied at week 0 was applied at week 14 to identify differences in post-induction expression associated with primary response. A larger proportion of the transcriptome was differentially expressed between responders and non-responders at week 14: 1364 genes for the infliximab-only comparison, 1544 genes for the adalimumab-only comparison, and 4841 genes pooling both drugs (Fig. 4.10). No significant interactions between drug and response were detected at the per-gene level. Given that sample sizes at week 0 and week 14 are comparable (Fig. 4.1), the overall signal-to-noise ratio is much stronger than at baseline.

After adjusting for cell composition, 1320/1364, 1515/1544, and 4653/4841 genes had dampened effects; and the numbers of significant genes dropped to 379, 177, and 1302; for the infliximab, adalimumab, and pooled analyses respectively. This again suggests many effects are mediated by differences in immune cell composition between responders and non-responders.

Modules including generic immune activation, monocytes, TLR and inflammatory signalling, and neutrophils were downregulated in responders; whereas B cell and plasma cell modules were upregulated (Fig. 4.11). These modules remained differentially expressed with the same direction of effect after adjusting for cell composition (Fig. 4.12), suggesting there is per-cell up or downregulation on top of abundance changes of the cell types expressing these modules. Modules related to antigen presentation (LI.M71, LI.M97.0, LI.M5.0), interferon (LI.M75, LI.M127, LI.M111.1), and dendritic cells (LI.M64, LI.M165) also appeared among significantly downregulated modules after cell composition adjustment. Directions of effect for the most significant modules were largely consistent between drugs, and there were few significant drug by response interaction effects. This is in contrast to the baseline responder vs. non-responder comparison, where many of the strongest effects in the pooled analysis were driven by stronger effects in one drug.

SIGLEC10 from the baseline analysis retained its significant association with primary response post-induction, with the same direction of effect (adjusted $\log_2 \text{FC} = 0.37$). Some genes previously

proposed as baseline markers of response in gut mucosa—*G0S2*, *TNFAIP6*, *S100A8*, and *S100A9* by Arijs *et al.* [363]; and *OSM* by West *et al.* [364]—were differentially expressed in post-induction blood in this study. The direction of effect for both sets of markers, downregulation in primary responders, also matches this study.

4.3.5 Magnification of expression changes from baseline to post-induction in responders

Given the stronger differences in expression between primary responders and non-responders at week 14 than week 0, I estimated the change in expression from week 0 to week 14 within the two groups, and also estimated the timepoint by response interaction. I performed only the pooled comparison to simplify the analysis, and because like the within week 14 comparison, change from week 0 to week 14 was relatively consistent between drugs, with exceptions noted.

Without adjusting for cell composition, 12 862 genes were differentially expressed in primary responders comparing week 14 vs. week 0, 8310 genes in primary non-responders, and 6320 genes had a significant interaction between responders and non-responders. After adjusting for cell composition, 5572 genes were differentially expressed in primary responders, 626 genes in primary non-responders, and 179 genes had a significant interaction. Of the genes differentially expressed between week 14 and week 0 in both primary responders and non-responders, and with a significant interaction between timepoint and response, nearly all (4885/4891 unadjusted for cell composition, 31/32 adjusted) were magnified by primary response, with the same genes having larger FCs in the same direction for primary responders (Fig. 4.13).

The most significant modules that changed from week 0 to week 14 in responders included upregulation of B cell (LI.M47.0), plasma cell (LI.M156.0), and T cell activation (LI.M7.1); and downregulation of immune activation (LI.M37.0), monocyte (LI.M11.0), neutrophil (LI.M37.1), and TLR and inflammatory signalling (LI.M16) modules (Fig. 4.14). Many of these are the same modules associated with response within the week 14 timepoint, with concordant directions of effect between the two analyses (Fig. 4.11), suggesting that a change in expression of these gene sets from week 0 to week 14 is what leads to the difference between responders and non-responders at week 14.

Adjusting for cell composition decreased the significance of a majority of modules (Fig. 4.15), especially for T cell modules in the adalimumab-only analysis. Magnification was also observed at the module level, with nearly all module effects aligned in the same direction in responders and non-responders, with significant interactions also in the same direction. In general, responders seem to experience greater changes in their gene expression from week 0 to week 14, presumably due to the drug.

4.3.6 Interferon modules with opposing expression changes in responders and non-responders

Fig. 4.13 also contains genes that were downregulated from week 0 to week 14 in responders, but upregulated in non-responders (“flipped”). At the module level, these flipped effects were apparent in the cell composition-adjusted analysis, for antiviral interferon signature (LI.M175),

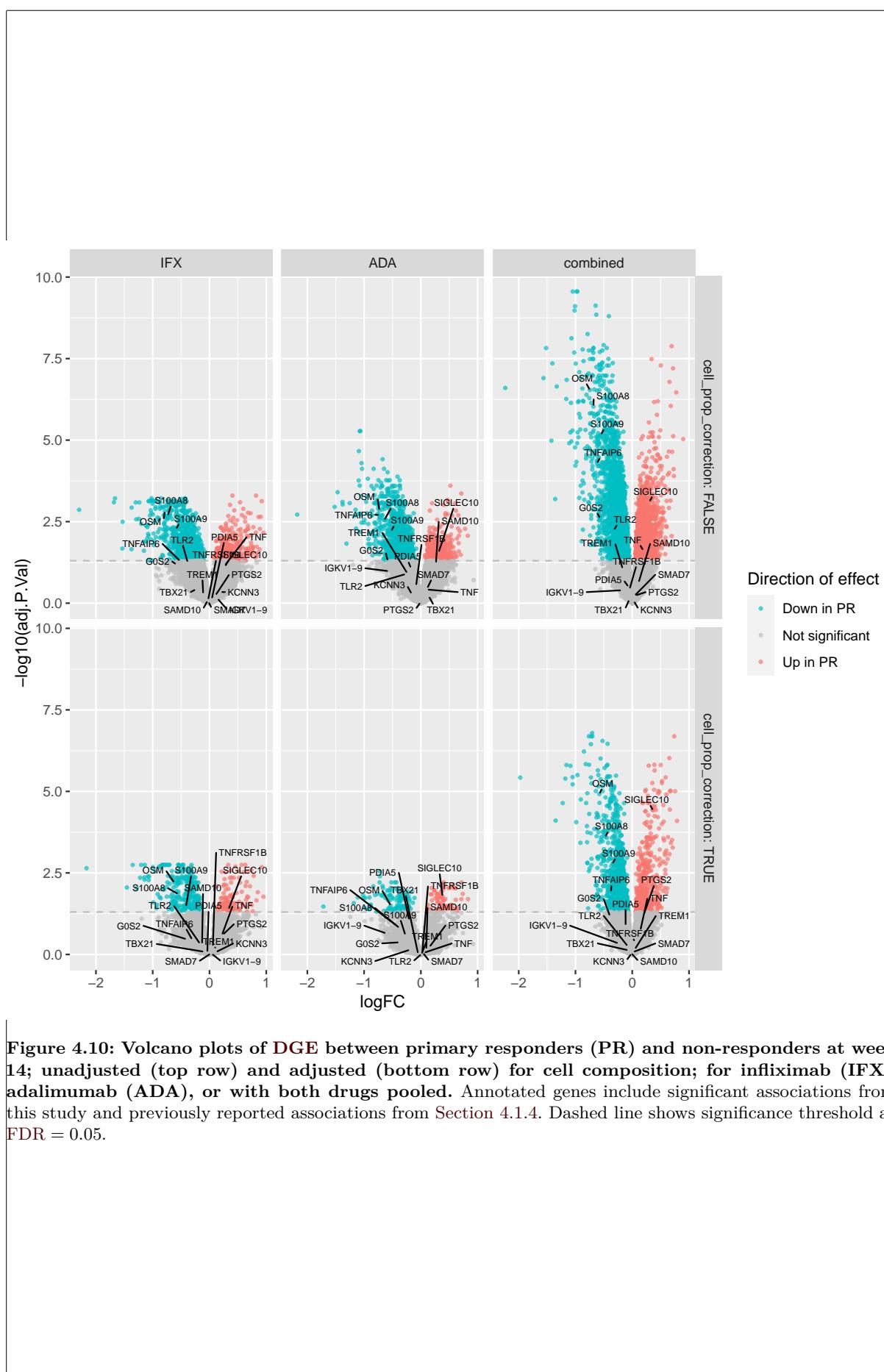
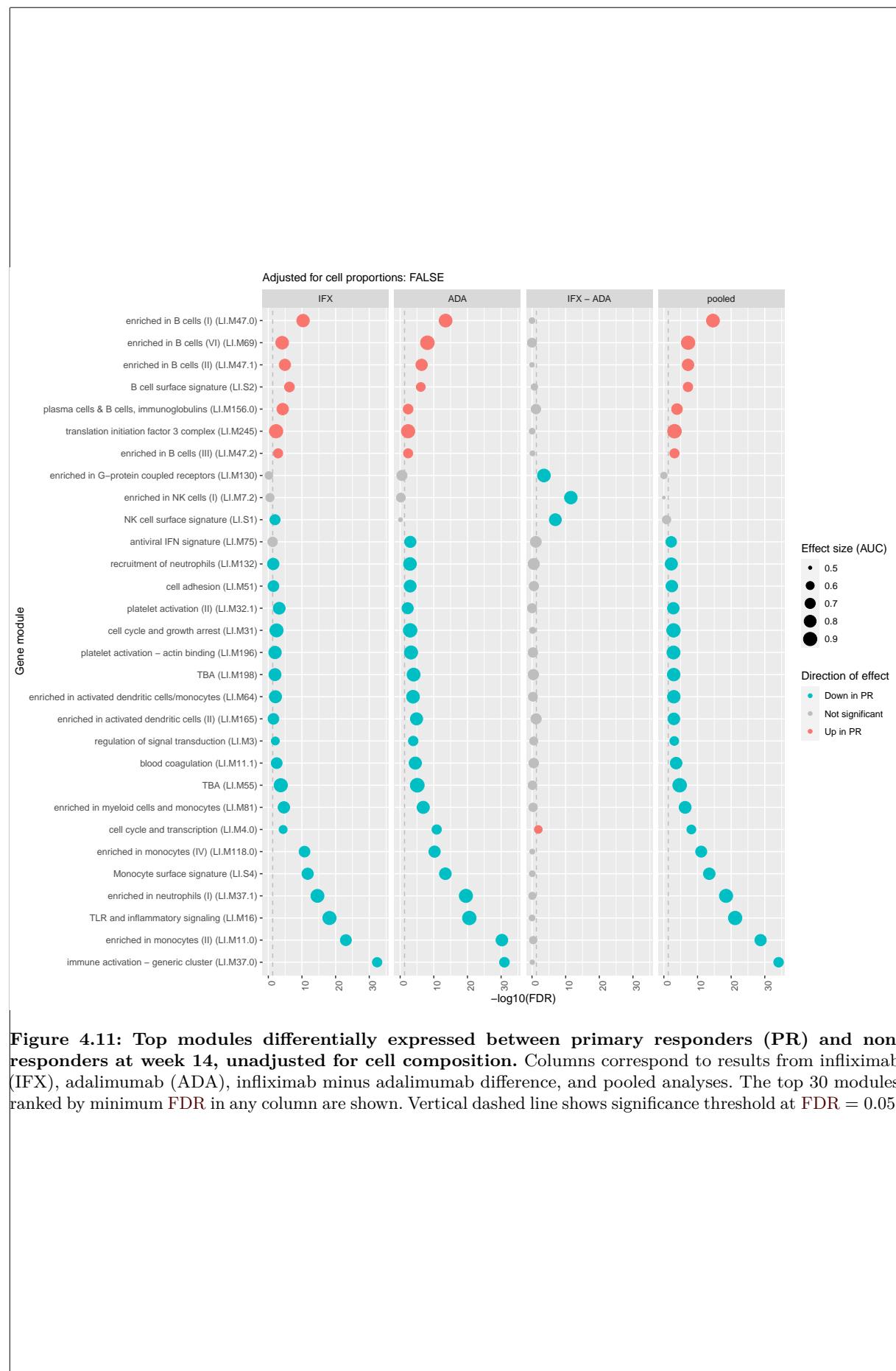


Figure 4.10: Volcano plots of DGE between primary responders (PR) and non-responders at week 14; unadjusted (top row) and adjusted (bottom row) for cell composition; for infliximab (IFX), adalimumab (ADA), or with both drugs pooled. Annotated genes include significant associations from this study and previously reported associations from Section 4.1.4. Dashed line shows significance threshold at FDR = 0.05.



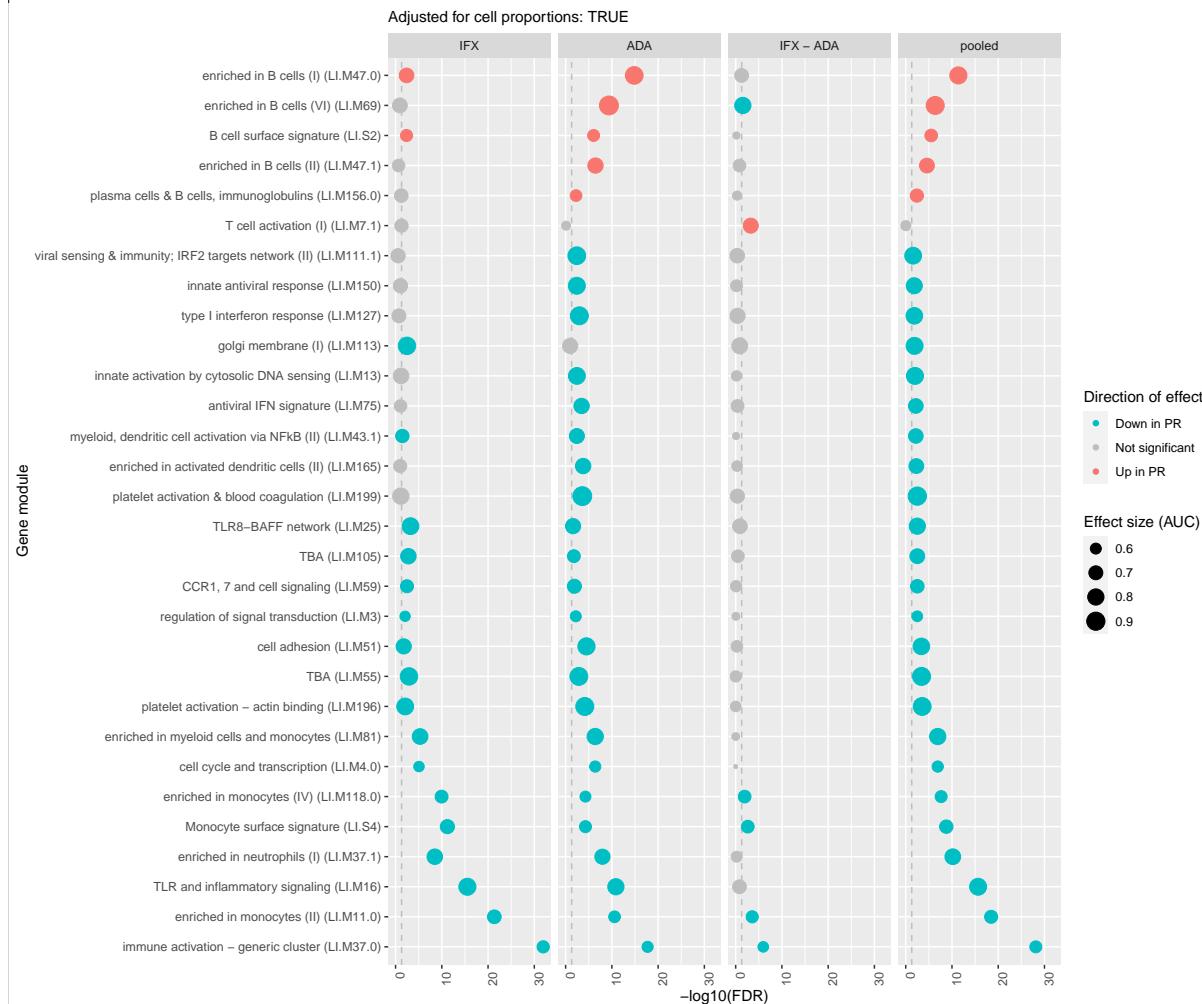


Figure 4.12: Top modules differentially expressed between primary responders (PR) and non-responders at week 14, adjusted for cell composition. Columns correspond to results from infliximab (IFX), adalimumab (ADA), infliximab minus adalimumab difference, and pooled analyses. The top 30 modules ranked by minimum FDR in any column are shown. Vertical dashed line shows significance threshold at FDR = 0.05.



Figure 4.13: Expression $\log_2 \text{FC}$ from week 0 to week 14 in primary responders (PR) versus non-responders (PNR), for genes that differentially expressed from week 0 to week 14 in both responders and non-responders, with a significantly different effect size between responders and non-responders. Results adjusted (right) and unadjusted (left) for cell proportions are shown. The identity line is shown by the dashed line. Most expression changes from week 0 to week 14 are magnified in primary responders, with a small proportion of changes in the opposite direction.

type I interferon response (LI.M127), and antigen presentation (LI.M95.0) modules (Fig. 4.15). I extended my gene set enrichment analyses to include modules from Chaussabel *et al.* [239] (prefixed “DC”); although these modules are on the whole poorly annotated compared to modules from Li *et al.* [240], interferon modules are well-annotated. *STAT2*, *GBP5*, and *PARP14* from Fig. 4.13 are annotated into an interferon module, DC.M3.4. *IFIT3* and *GBP2* are also annotated into separate interferon modules, DC.M1.2 and DC.M5.12. Adjusted for cell composition, these modules were all significantly upregulated at week 14 in non-responders only (DC.M3.4, $FDR = 3.45 \times 10^{-21}$; DC.M1.2, $FDR = 9.49 \times 10^{-16}$; DC.M5.12, $FDR = 1.36 \times 10^{-13}$; Fig. 4.16).

4.3.7 Sustained expression differences between primary responders and non-responders during maintenance

As PANTS is an observational study, it was able to include patients who continued with anti-TNF therapy even after meeting the definition of primary non-response at week 14. For both responders and non-responders, expression data was also available from blood samples around week 30 and week 54, and at additional visits scheduled in the event of secondary LOR. To test for general differences in expression over time between responders and non-responders, I fit a natural cubic spline to the expression of each gene as a function of study day. This analysis was performed only with drugs pooled due to lower sample sizes at later timepoints.

Without adjusting for cell composition, 4426 genes were differentially expressed between responders and non-responders; 210 genes were differentially expressed after adjustment. To identify distinct trajectories of expression over time, I hierarchically clustered those 210 genes by their mean expression in responders and non-responders at each timepoint, and determined the

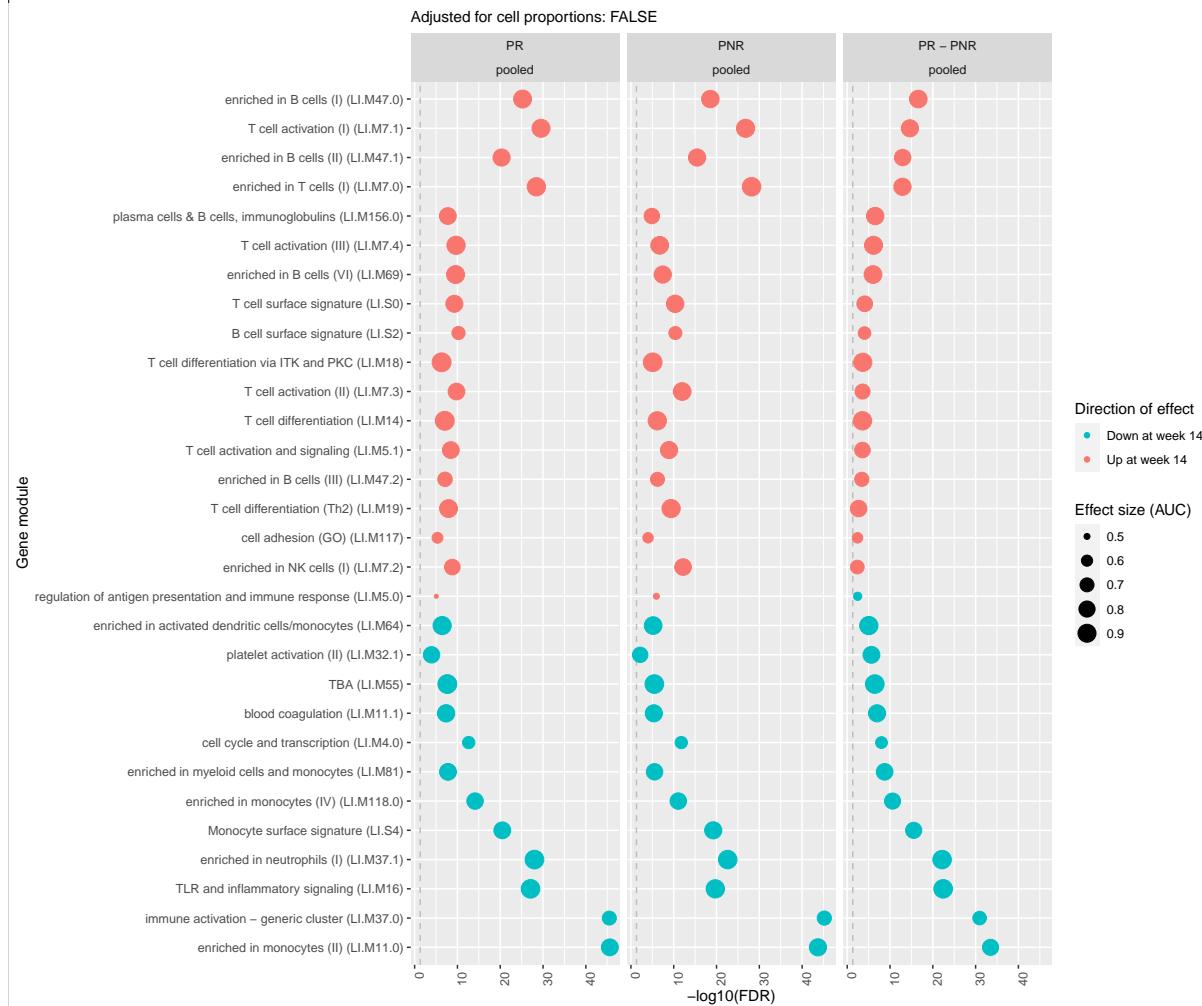
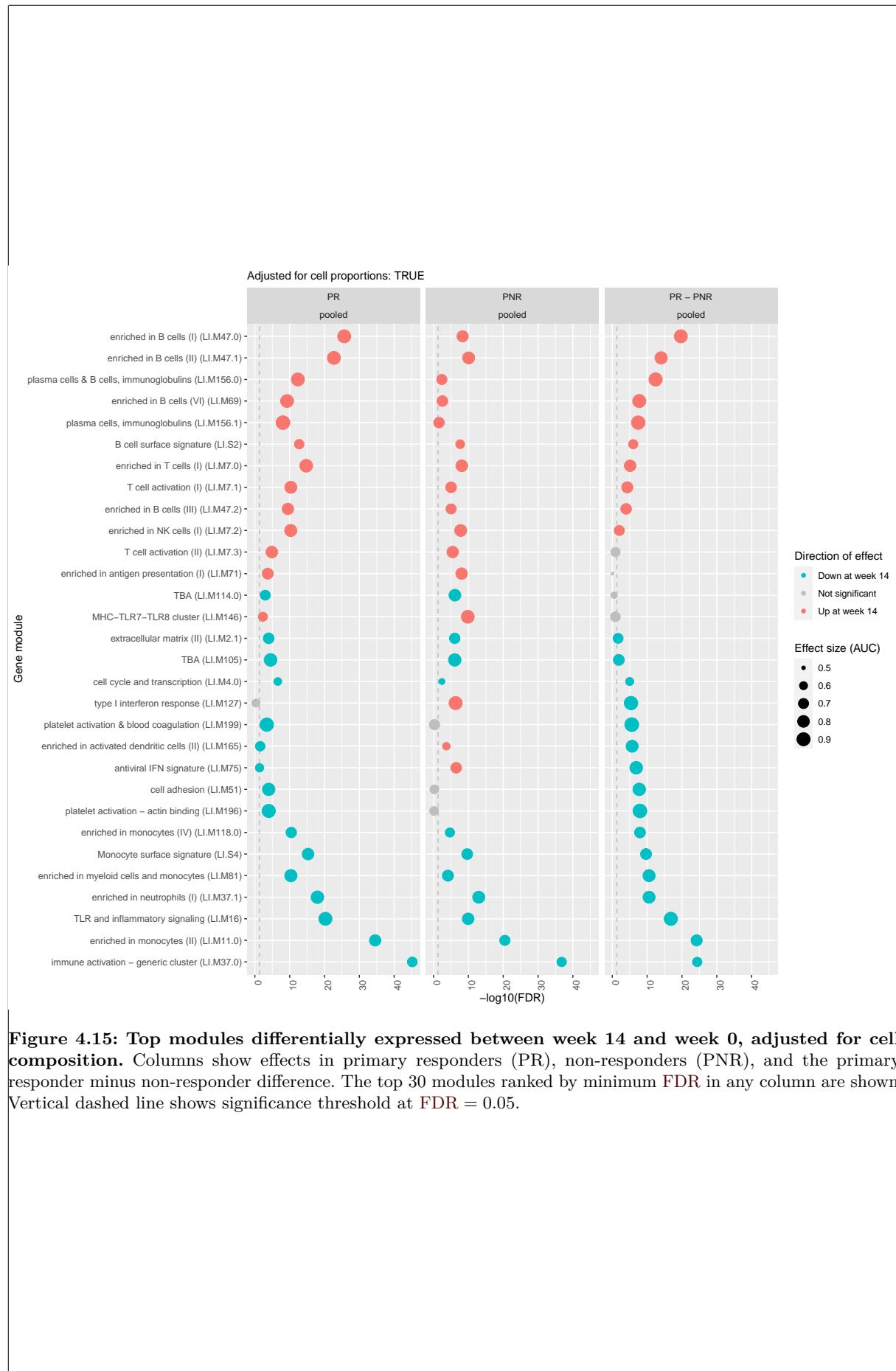


Figure 4.14: Top modules differentially expressed between week 14 and week 0, unadjusted for cell composition. Columns show effects in primary responders (PR), non-responders (PNR), and the primary responder minus non-responder difference. The top 30 modules ranked by minimum FDR in any column are shown. Vertical dashed line shows significance threshold at FDR = 0.05.



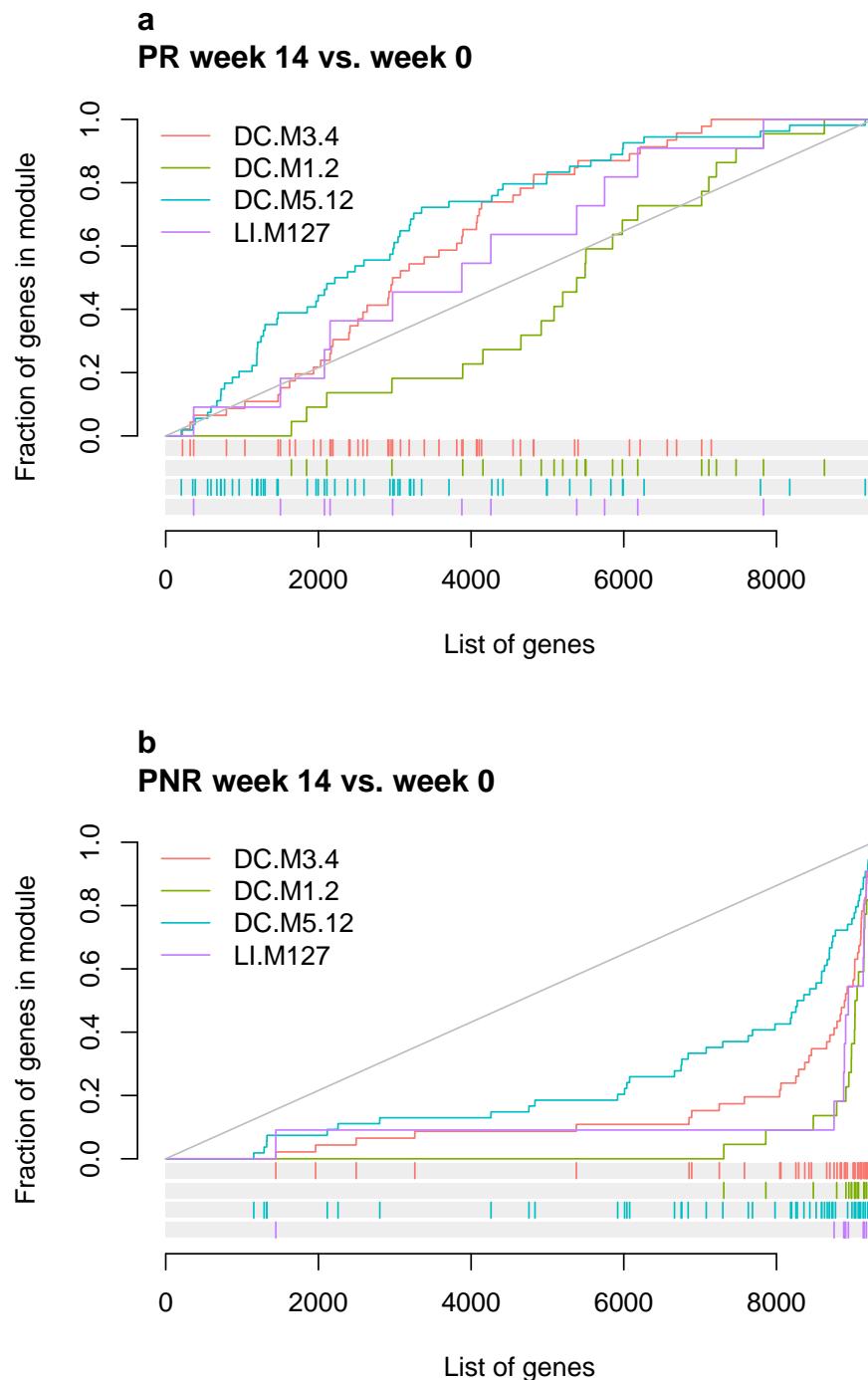


Figure 4.16: tmod evidence plots showing interferon-related modules specifically upregulated from week 0 to week 14 in primary non-responders (PNR), but not in primary responders (PR). Genes were ranked in ascending order by week 14 versus week 0 DGE z-statistic. The ranks of genes in interferon-related modules are indicated by colored rug plots. Colored curves show the cumulative fraction of genes in each module. For non-responders, these modules are enriched for large ranks (large, positive z-statistics). The area under the colored curves are the effect sizes (area under the curves (AUCs)). The null of randomly-distributed ranks is shown by the grey diagonal line.

optimal number of clusters by the gap statistic method (Fig. 4.17). Six distinct clusters were proposed (Fig. 4.18). Many of the 210 genes had previously been identified as having significant differences in expression between responders and non-responders either within week 14, or for change in expression from week 0 to week 14. Cluster 1 contained mainly previously identified genes (Fig. 4.19); and was enriched for modules such as myeloid cells and monocytes (LI.M81, hypergeometric test, **FDR** = 2.11×10^{-6}), platelet activation (LI.M196, **FDR** = 1.35×10^{-5}), immune activation (LI.M37.0, **FDR** = 1.44×10^{-4}), and TLR and inflammatory signalling (LI.M16, **FDR** = 2.36×10^{-3}). The spline analysis highlights that expression differences at week 14 are maintained at week 30 and week 54.

The highest proportions of genes uniquely identified as significant by the spline analysis were in cluster 2 (26/31) and cluster 4 (15/20). Cluster 2 was enriched in Li *et al.* [240] B cell modules (LI.M47.0, **FDR** = 1.53×10^{-6} ; LI.M47.1, **FDR** = 4.53×10^{-5}) previously identified as having a greater increase from week 0 to week 14 in primary responders than in primary non-responders (Fig. 4.15), matching the observed cluster trajectory. Cluster 4 was not enriched in any modules from Li *et al.* [240], but was enriched for a B cell module (DC.M4.10, **FDR** = 1.37×10^{-3}) from Chaussabel *et al.* [239]. Although no genes were significantly associated with response at week 0 (Fig. 4.7), the genes in cluster 4 were coordinately downregulated as a set in primary responders (CERNO test, $p = 6.18 \times 10^{-25}$).

Cluster 3 was enriched for type I interferon response (LI.M127, **FDR** = 0.01) and interferon (DC.M3.4, **FDR** = 5.27×10^{-4}) modules, as well as genes that contain putative transcription factor (TF) binding motifs for interferon regulatory factors *IRF7* (g:Profiler [313] term ID TF:M00453_1, adj. $p = 0.01$) and *IRF8* (TF:M11684_1, adj. $p = 0.01$; TF:M11685_1, adj. $p = 0.01$). The cluster trajectory shows direction of expression change is opposing in responders and non-responders from week 0 to week 14, followed by sustained differences at week 30 and week 54. The trajectory and interferon-related gene set enrichments are consistent with those identified in Section 4.3.6. Of the nine genes in this cluster, eight genes (*STAT1*, *BATF2*, *GBP1*, *GBP5*, *IRF1*, *TAP1*, *APOL1*, *APOL2*) had significant interaction between week 0 to week 14 expression change and response status, whether or not correcting for cell composition. However, only *GBP5* was differentially expressed from week 0 to week 14 in both responders and non-responders, and only when unadjusted for cell composition (Fig. 4.13). This indicates that small and opposite effects in responders and non-responders at the gene level are best detected in the interaction analysis that explicitly tests the difference, and in the spline analysis with the support of additional data from week 30 and week 54.

4.3.8 Limited evidence for changes in genetic architecture of gene expression over time

Given the substantial changes in expression from baseline to post-induction after starting the drug, and the differing trajectories observed in responders and non-responders, I performed eQTL mapping to identify common genetic variants associated with expression that may contribute to these differences. Variants *cis* (within 1 Mb of the transcription start site (TSS)) to 15 040 genes were tested for association. Mapping was performed within each timepoint (weeks 0, 14, 30, and 54), followed by joint analysis of per-timepoint eQTL summary statistics and control for

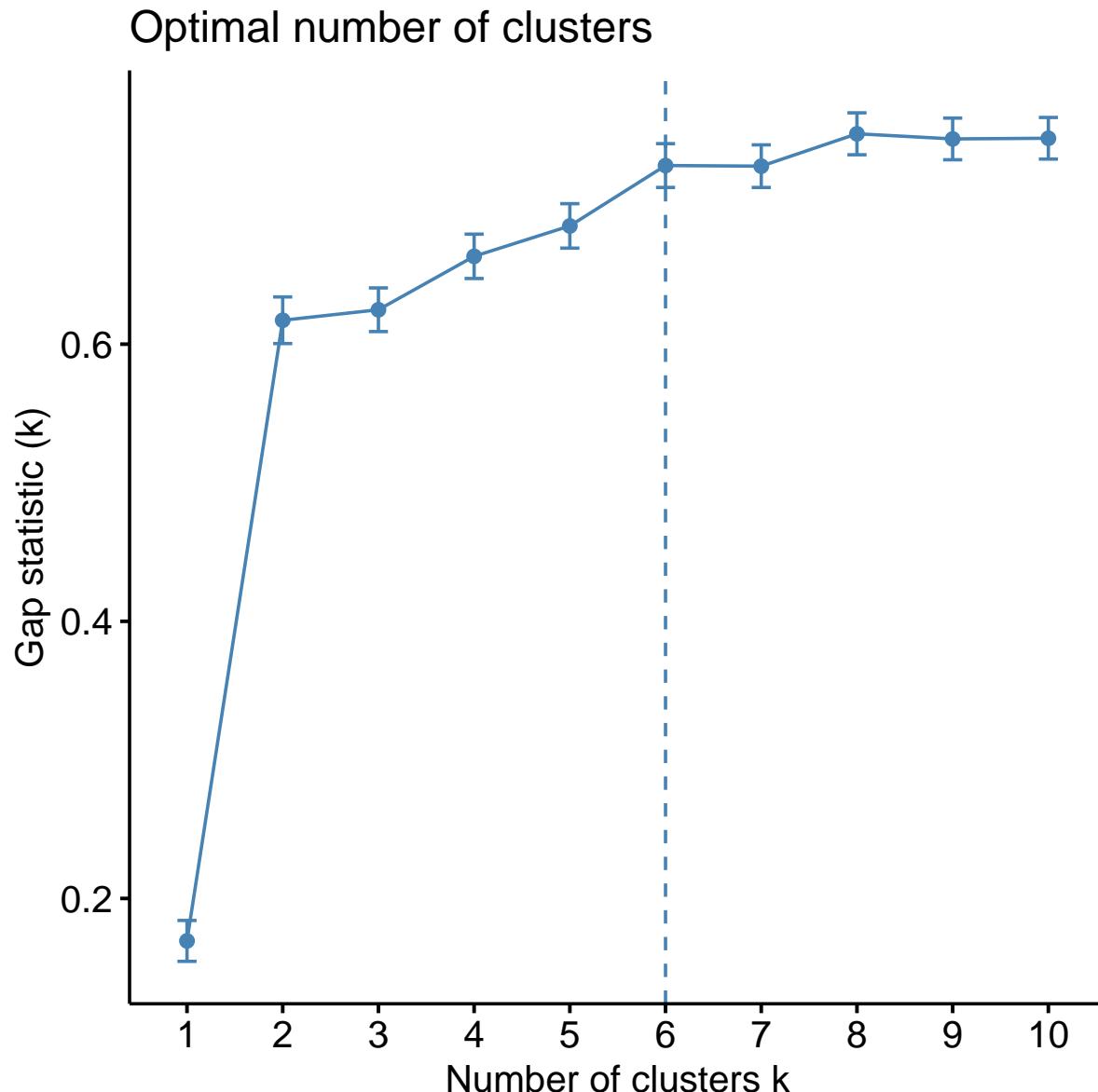


Figure 4.17: Gap statistic versus cluster number k . Error bars derived from 500 bootstraps. Here, the optimal number was $k = 6$ by the `factoextra::fviz_nbclust firstSEmax` criteria (<https://rpkgs.datanovia.com/factoextra/index.html>).

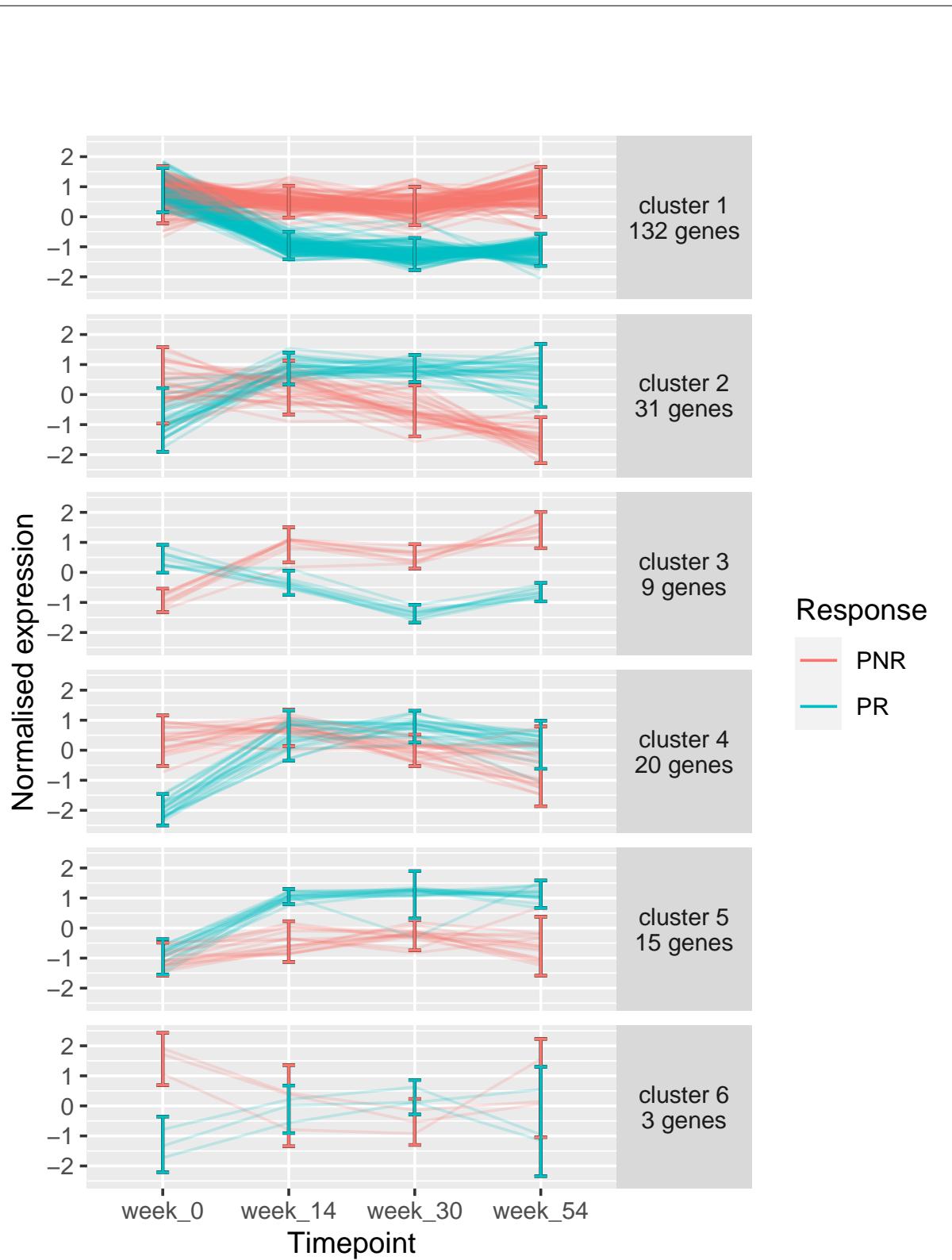


Figure 4.18: Normalised expression over the timepoints for genes in the six identified clusters. Log scale expression for each gene was standardised before clustering. Error bars show the mean \pm standard deviation for the genes at each timepoint in primary responders (PR) and non-responders (PNR).

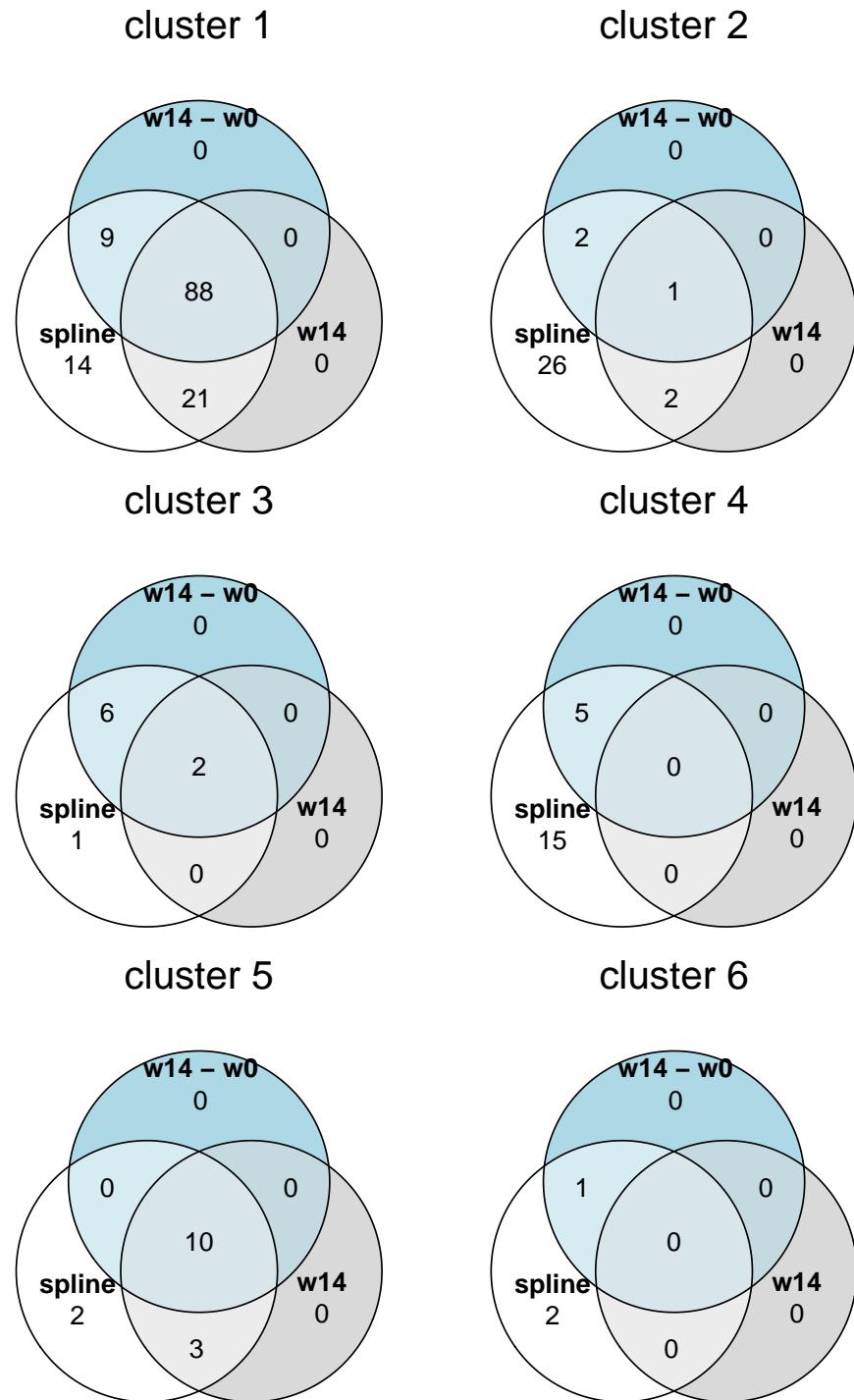


Figure 4.19: Overlap of genes differentially expressed between responders and non-responders from the spline model, the week 14 responder versus non-responder contrast (w_{14}), and the interaction between week 0 to week 14 change and response status contrast ($w_{14} - w_0$). Clusters 2 and 4 largely contained genes identified as significant only by the spline model.

multiple testing using `mashr` [281].

The majority 11 156/15 040 (74.2 %) of genes were eGenes (a gene with at least one significant *cis*-eQTL) in at least 1 timepoint (**LFSR** < 0.05). The variant with the lowest **LFSR** over all timepoints for each gene was chosen as the lead variant (eSNP) for that gene. Most eSNPs were significant at multiple timepoints: 999 at one timepoint, 381 at two timepoints, 526 at three timepoints, and 9250 at all four timepoints. I compared eSNP effect sizes between week 0 and each of weeks 14, 30, and 54 to identify reQTLs with a significant difference in effect versus baseline, as they may explain changes in expression from baseline. Most eSNPs were shared across timepoints; only six eSNP-eGene pairs were significant reQTLs (difference in betas BH FDR < 0.05): 1/6 between week 30 and week 0, and 5/6 between week 54 and week 0 (Fig. 4.20). Of the six eGenes with reQTLs between week 54 and week 0, *NMI* and *EPSTI1* both have magnified eQTL effect sizes at week 54 compared to week 0, and both are annotated to contain putative binding motifs for *IRF8* and *IRF2* (g:Profiler term IDs TF:M11685_1 and TF:M11665_1). However, direct interpretation of these reQTLs is complicated by changing cell composition in bulk expression data (discussed in Section 3.2.10 and Section 5.3).

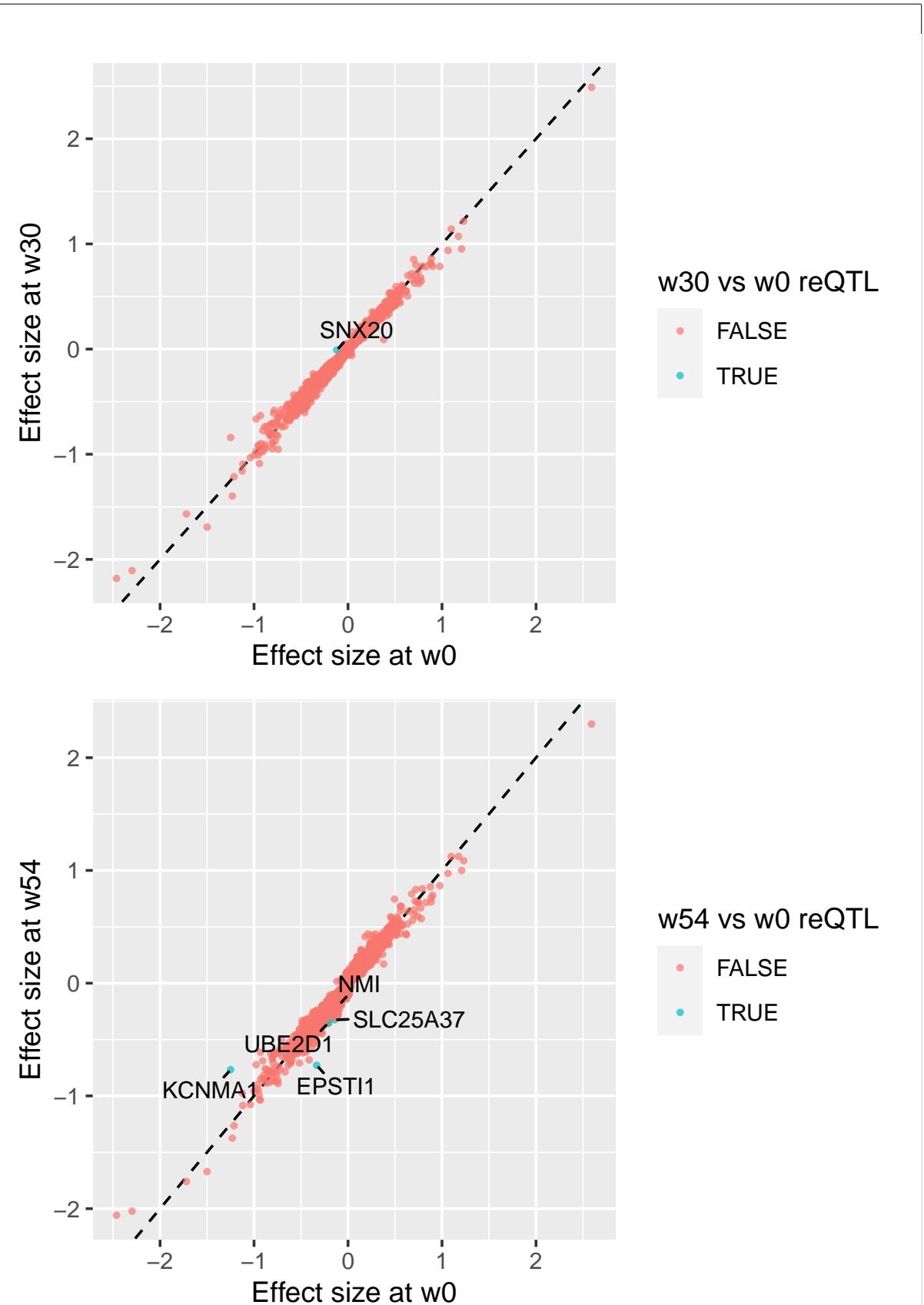


Figure 4.20: Post-treatment eQTL effect sizes at week 30 and week 54 compared to baseline effect sizes at week 0. Significant reQTLs at BH FDR < 0.05 are labelled.

4.4 Discussion

In PANTS, a cohort of CD patients receiving infliximab or adalimumab anti-TNF therapy for the first time, there were substantial differences in whole blood gene expression between primary responders and non-responders. At baseline, the greatest differences in expression were observed between future responders and non-responders to infliximab, with increased expression of monocyte, neutrophil, and dendritic cell gene modules in responders, and decreased expression of T cell and NK cell modules. These effects appear to be infliximab-specific, and are attenuated after adjusting for the proportions of six major immune cell types, suggesting expression differences may be driven by mediation via the proportions of these cell types.

In contrast, future responders to adalimumab had lower baseline expression of plasma cell and cell division modules. The three hits from the gene-level adalimumab-only analysis implicate similar cell types; *IGKV1-9* encodes the immunoglobulin light chain variable region that forms part of antibodies produced by plasma cells, *KCNN3* is annotated to a plasma cell surface signature module (LI.S3 [240]), and the expression of both *KCNN3* and *PDIA5* are correlated with blood plasmablast frequencies [161]. Gaujoux *et al.* [366] observed lower baseline plasma cell abundances in infliximab responders than in non-responders, and hypothesised that plasma cell survival is supported by increased TNF levels in non-responders. Plasma cells also formed part of a correlated module of cell populations identified by Martin *et al.* [365], where lower module expression was associated with better response to anti-TNFs in a cohort with both infliximab and adalimumab patients. However, both these studies were conducted in gut biopsy samples, and there was no mention of strong between-drug heterogeneity.

The adalimumab-specific associations were more significant after cell proportion adjustment, which may indicate per-cell downregulation rather than cell abundance being associated with response. However, cell composition differences mediated by rarer cell types that have abundances poorly captured by the six major types used in the model will be poorly adjusted for. For example, plasma cell proportions are only weakly correlated with other cell types in the healthy immune system [388], although the relationship may differ in CD patients. It has also been shown that DGE analyses with correction for only common blood cell types can identify associations that are proxies for rare cell types [389]. If this is the case, the role of the cell composition estimates for adalimumab-specific effects may be more akin to precision variables, which would be consistent with increased significance after adjusting.

The between-drug heterogeneity for baseline associations is puzzling, especially the greater effect of cell composition adjustment for the infliximab-only model. Baseline patient differences between drugs could offer a partial explanation. There may be characteristics not listed in Table 4.1 that differ between patients on different drugs [355]. In the full PANTS cohort, lower albumin, higher CRP, and higher faecal calprotectin levels in infliximab patients suggest that they may have had greater disease severity. Differences may be driven by patient or physician preference, for example, patients with more severe disease are often given infliximab rather than adalimumab*. I have not yet been able to access clinical variables such as CRP and faecal calprotectin levels, variables that one could consider adjusting for in the DGE models. A richer

* Kennedy, N. A., personal communication, 4 June (2020).

phenotype dataset containing clinical variables has been requested from collaborators.

The strongest single-gene association in the pooled analysis was *SIGLEC10*, which had reduced significance and a comparable effect size post-adjustment, where baseline expression was approximately 25 % higher in responders. Direction of effect was consistent between drugs, although the association was most significant in infliximab without cell composition adjustment. In IBD, small molecules called damage-associated molecular patterns (DAMPs) are released due to tissue damage and cell death, and further promote inflammation through pathogen-sensing pattern recognition receptor (PRR) pathways that include toll-like receptor (TLR) family receptors [339, 390]. For instance, faecal calprotectin, a marker for IBD activity, is a complex of two DAMPs, S100A8 and S100A9 [339]. *SIGLEC10* has been shown to repress DAMP-mediated inflammation through binding CD24 [390]. *SIGLEC10* is expressed on B cells, monocytes, and eosinophils [391]. Of these cell types, module level results posit monocytes as the most likely candidate cell type to be increased in anti-TNF responders. In monocytes, *SIGLEC10* gene expression is more specific to the CD16⁺ monocytes [392], and in particular the CD14⁺CD16⁺⁺ non-classical monocytes rather than the classical CD14⁺⁺CD16⁻ or intermediate CD14⁺⁺CD16⁺ subsets [393]. In PANTS, it was suggested by Kennedy *et al.* [355] that higher inflammatory load, as indicated by low baseline albumin levels, may result in low week 14 drug levels due to faster drug clearance. Low drug levels at week 14 were in turn associated with non-response. A hypothetical model might be high baseline *SIGLEC10* expression reflecting higher proportions of CD16⁺ monocytes (or lower proportions of CD16⁻ monocytes), decreased DAMP-mediated inflammation and increasing chance of primary response, possibly by affecting drug clearance rate. This is an extremely tentative model: both the cell proportion estimates and module definitions used thus far only represent monocytes as a whole, lacking the resolution to properly explore shifts in monocyte subsets. It may be possible to use expression of monocyte subset marker genes, such as those identified by Villani *et al.* [393], to improve the resolution of the cell proportion estimates.

Despite the strong heterogeneity in effects between drugs, consistent module-level effects that emerged after adjusting for cell composition included baseline upregulation of MHC-TLR7-TLR8, antigen presentation, and interferon modules in responders. As mentioned above, TLR receptors are involved in pathogen sensing, and TLR7 and TLR8 are endosomal proteins primarily expressed in monocytes, macrophages, and dendritic cells (DCs), part of an antigen presentation pathway that senses bacterial DNA and activates downstream innate immune pathways including type I interferon response [318]. Type I interferons have pathogenic or protective roles in many IMIDs [394]. It has been suggested that type I interferon responses induced via TLR7 and TLR8 can suppress colitis in mouse models, and play a role in maintaining gut homeostasis [395, 396], so upregulation here may again represent a less severe baseline disease in future responders.

Most previously reported baseline markers in blood and gut biopsies were non-significant in this study. For gut markers, this may not be unexpected. Although a subset of gut infiltrating immune cells and their precursors may also be circulating, genes specific to epithelium and immune cell types that differentiate after they migrate into tissues (e.g. monocyte-derived macrophages) are difficult to observe in blood. For blood markers, I sought to clarify the conflicting results in the literature about the association of *TREM1* expression in blood with anti-TNF response

[366, 367]. *TREM1* is expressed in myeloid lineage cells such as monocytes and macrophages; Villani *et al.* [393] reported that *TREM1* expression is most specific to classical monocytes and a newly identified subtype within the intermediate monocytes (“Mono3”). I did not find *TREM1* to be significantly differentially expressed in PANTS. The direction of effect corresponded to increased expression in responders, matching the Gaujoux *et al.* [366] direction of effect in blood. The strongest effect was observed in the infliximab-specific comparison without cell proportion adjustment, which may be another indication that baseline monocyte cell proportions are associated with response. There are many factors that could explain failures to replicate reported markers or identification of different markers from study to study. Many existing studies pool cohorts with different anti-TNF drugs due to the scarcity of large datasets, yet even within the PANTS cohort, there appears to be heterogeneity between drugs. There are between-study differences in the definition of primary response, such as endoscopic healing [366] versus scoring on clinical parameters [367]. Any two studies are unlikely to have adjusted for the same combinations of covariates in modelling, and some covariates like cell composition are very influential for bulk expression data. Finally, small sample sizes have considerable sampling error. Set-based associations that draw on changes in multiple genes, such as the module associations from this study, may be more reproducible compared to single-gene markers. A future aim will be to see if the identified baseline module associations also imply that response status can be predicted from baseline expression. Because modules associated with response appear to be mediated by cell proportions, much of the predictive ability may also lie in differences in cell proportions between responders and non-responders. Indeed, Gaujoux *et al.* [366] noted that adjusting expression for cell composition resulted in gut gene signatures that were worse at discriminating responders from non-responders. Testing the abundances of specific subpopulations for association with response (e.g. CD16⁺ monocytes or plasma cells) can also be viewed as a type of set-based test that represents a set of cell-type specific genes, and thus may also be more reproducible than single-gene markers.

A much larger proportion of the transcriptome was associated with response after the induction period at week 14. Module associations showed downregulation of immune activation, TLR, inflammatory, monocyte, and neutrophil modules in responders; and upregulation of B and T cell modules. Similar module associations were also found when considering modules differentially expressed from week 0 to week 14. The differences between responders and non-responders at week 14 were qualitatively similar to the differences between week 14 and week 0 in responders, suggesting there may be relatively little change in the transcriptome of non-responders after anti-TNF induction. Associations were generally consistent between drugs for both the within week 14 and change from week 0 to week 14 analyses, perhaps because any baseline transcriptomic differences between patients taking different drugs were diluted by the large transcriptomic perturbation caused by taking an anti-TNF drug. Many of the same modules were also significant regardless of cell proportion correction. A general reduction in immune activation in responders at week 14 is presumably due to successful inhibition of TNF-mediated inflammation by the anti-TNF drug. Decreased inflammation correlates with reduced neutrophil activation and reduced monocyte recruitment [397], supported by the observed downregulation of neutrophil and monocyte modules. Apoptosis of monocytes induced by anti-TNF in CD patients

has also been previously described [398]. Certain B cell subsets are reduced in the blood of IBD patients compared to controls [399], so upregulation of B cell modules at week 14 may represent a shift towards health. Another potential explanation would be increased immunogenicity due to higher drug levels in responders [355], although lack of between-drug heterogeneity for the B cell signal is not consistent with the greater immunogenicity of infliximab. Overall, it is difficult to glean exact mechanisms from an observational study design, with bulk expression data, and using such broad module definitions.

Some previously identified baseline gut markers of response that were not differentially expressed in blood at week 0, were differentially expressed at week 14. *S100A8* and *S100A9*, identified as markers by Arijs *et al.* [363], which encode components of the inflammatory marker CRP, were downregulated in week 14 responders. The cytokine *OSM*, which promotes inflammation in gut stromal cells [364], was similarly downregulated. Although it is pointless to use a week 14 marker to predict a response that is defined at week 14, this does demonstrate that gut markers can coincide with blood markers if expressed in immune cells present in both tissues.

When considering the interaction between change from week 0 to week 14 and response, the general pattern is magnification in responders, where the same expression changes occur with greater magnitude than in non-responders. A potential hypothesis is a continuum of response from non-response to response. Gaujoux *et al.* [366] found changes in cell proportions in response to anti-TNF treatment were magnified in responders, also supporting response as continuous phenotype. This study demonstrates a similar trend at the transcriptional level. There were some rare exceptions to magnification for genes and modules in the type I interferon pathway. These showed upregulation in non-responders from week 0 to week 14, yet were either downregulated or not significantly different in responders. Single-gene examples include the interferon-induced guanylate-binding proteins *GBP2* and *GBP5* [400], and *STAT2*, a key transcription factor for interferon-stimulated genes [244]. Genes such as *IFIT3* and *STAT2* are more strongly induced by type I interferons compared to type II [401]. A study of RA, an IMID also treated with anti-TNF drugs, likewise found increases in type I interferon-regulated gene expression in blood after infliximab treatment to be associated with poor clinical response [402].

A spline model of expression over all four timepoints confirmed the above observations made in week 0 and week 14 samples. Two main clusters of genes (clusters 1 and 5) contained mostly genes significantly associated with response in the two pairwise comparisons: within week 14, and change from week 0 to week 14. An example is the most significant single-gene association from cluster 1 in spline model, *KREMEN1*, which is also one of most significant associations in the pairwise comparisons. *KREMEN1* is part of an inflammatory apoptotic pathway in gut epithelium [403], and is downregulated in responders post-induction. The trajectories of expression for genes in clusters 1 and 5 confirmed that changes in expression post-induction were generally greater for responders, and in addition demonstrates that post-induction expression differences between responders and non-responders are sustained in samples taken around week 30 and week 54 during the anti-TNF maintenance period. In PANTS, Kennedy *et al.* [355] found that “continuing standard dosing regimens after primary non-response was rarely helpful” for inducing remission by week 54. This phenomenon may have a transcriptomic basis, although non-responders in the PANTS RNA-seq data were selected to exclude patients in remission by week 54, so trajectories

for non-responders at week 14 that eventually achieved remission could not be observed.

Making use of data from later timepoints allowed more subtle effects to be detected in the spline analysis. Clusters 2 and 4 were enriched for B cell genes that were not significantly different at the gene-level in the within week 0 comparison, although some downregulation of B cell and plasma cell modules was detected. Cluster 3 reproduced the observation that interferon-induced genes have opposing trajectories of expression in responders and non-responders. Expression of these genes was higher in responders at week 0 and lower at all post-treatment timepoints. The cluster contains genes such as *STAT1*, *IRF1*, and *TAP1* that are induced by both type I and type II interferons [401]. I propose that blood expression of interferon-related genes is an attractive target for future studies of the biological basis of anti-TNF response, and for use in building predictive models of primary response status. Since the difference is maintained until week 54, by which time patients would have received many doses of drug, it is more likely that response is due to some biological property of an individual patient. Studies of anti-TNF response in RA patients have also found high baseline interferon activity in blood to be associated with good clinical response [404, 405]. It should be noted that the number of clusters is only the optimal number determined in this dataset, and does not imply that genes in different clusters represent biologically distinct pathways. Clusters 2 and 4 have similar trajectories and enrichments for B cell genes, and interferon pathway genes appear in both clusters 1 and 3.

Finally, I attempted to determine if there were changes in genetic architecture of expression over time, which could indicate that expression response to anti-TNF has a genetic component. Out of all significant lead eQTLs for 11 156 genes, only six reQTLs were detected between baseline and any one of the three post-treatment timepoints. Although no enrichment analyses were attempted due to the small number of associations, *NMI* and *EPSTI1* are both interferon-induced genes with significant reQTLs that had their strongest effect size on expression at week 54. Given the issues with doing a reQTL analysis in bulk expression data are similar to those encountered in Chapter 3, I did not place emphasis on interpreting these small numbers of associations. I would also like to verify that these significant reQTLs are not artifacts from shrinkage of effect sizes in the joint eQTL model, as their posterior effect sizes from `mashr` were very different from the input effect sizes from the per-timepoint models. If these hits are indeed reproducible by complementary methods such as allele-specific expression (ASE) [406], it may then be worth introducing genotype-response interaction terms into the eQTL models to identify eQTLs with differing effects in responders and non-responders. Given there is prior interest in the interferon pathway from DGE analyses, a more statistically powerful approach may be to generate a continuous interferon pathway score for each sample, which would then act as the interacting variable, similar to the approach of Davenport *et al.* [96].

Several threats to the validity of the study remain to be discussed. The most pressing may be the meaning of time in the study. For pairwise DGE comparisons, expression trajectory clustering, and reQTL mapping, samples were divided into four discrete timepoints that corresponded to the major visits in PANTS; whereas the DGE spline model was fit to study day directly. Study day has substantial variation around the target for later timepoints. The particular target timings for post-baseline visits (weeks 14, 30, 54) were chosen so that patients on infliximab (8 weeks between doses) and adalimumab (2 weeks between doses) could both be sampled with the same

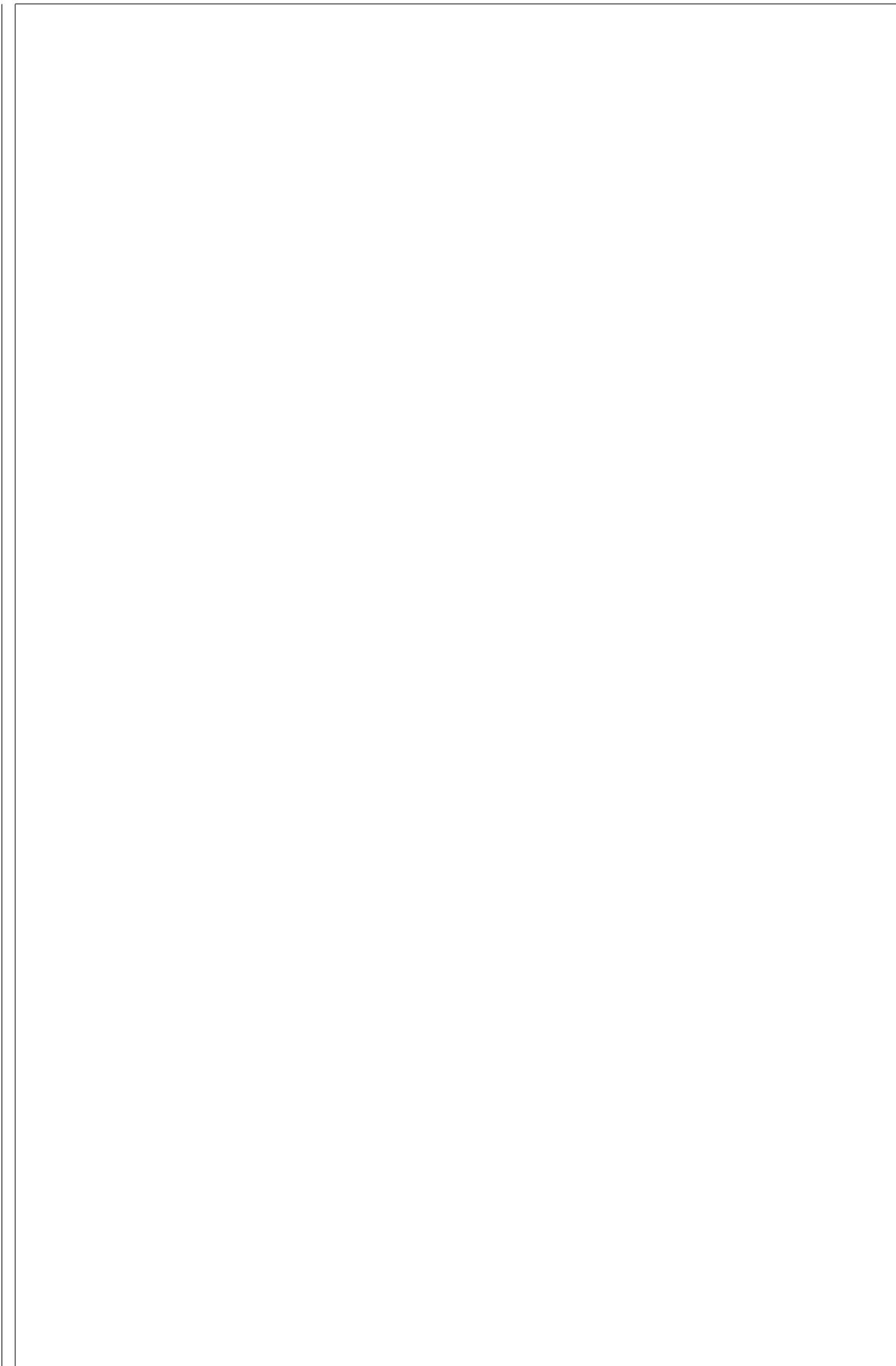
visit structure. Drug levels peak sharply after each dose and decline exponentially over time. To capture trough drug levels, visits were scheduled to be as close as possible to the next scheduled drug dose (within a week) Neither modelling approach is perfect; matching patients by timepoint and study day are merely attempts to gather samples matched by trough drug level.

A further complication is the inclusion of LOR samples in analyses. One treatment option after LOR is dose escalation, which may raise trough drug levels for all subsequent visits for those patients. However, since the PANTS protocol allows for LOR visits that coincide with major visits to be labelled as a major visit, there is no guarantee that simply excluding samples labelled as LOR would resolve this. The best solution may to explicitly model measured serum drug levels as a covariate, where like cell proportions, it would likely act as a mediator of some associations with response. I did not do this as data missingness would reduce the sample size by about 40 % in this study. Finding a suitable normalisation of drug level for use in pooled drug analyses would also be challenging. Infliximab and adalimumab have differing pharmacokinetics; infliximab has higher peak concentrations, higher peak-trough ratios, and shorter half-life. The same serum concentrations of infliximab and adalimumab also have different biological effects due to differing therapeutic windows [350, 407, 408].

The effect of differential drop out in responders and non-responders has not been explored. There are three main mechanisms of missing data: missing completely at random (MCAR), where probability of data being missing is independent of both observed and missing data; missing at random (MAR), where probability of data being missing conditional on observed data is independent of missing data; and missing not at random (MNAR), where probability of data being missing depends on missing data [409]. Even conditional on response status, it is more likely that expression data from more extreme non-responders is missing for later timepoints, so the likely mechanism here is MNAR, thus the linear mixed models used in this study may be biased. If it is indeed the most extreme non-responders dropping out, the estimation of responder versus non-responder effects may be conservative in the spline analysis. Note there is no sidestepping a MNAR mechanism by analysing only the complete cases, since they will differ systematically from the sample as a whole [410].

In conclusion, it remains unclear whether there are any robust single-gene expression markers for anti-TNF response in the whole blood of CD patients at baseline. Baseline module associations were observed, but there was unexpected heterogeneity between infliximab and adalimumab patients, so it remains to be seen if such associations will be replicated in other cohorts. Large upcoming datasets with drug response phenotypes such as the 1000IBD project [411] will be invaluable for attempted replication of the associations found in PANTS. Expression differences between responders and non-responders were more distinct at timepoints after the induction period. I found type I interferon genes were more upregulated post-treatment in non-responders, going against the general trend of magnified transcriptomic change in responders. Given that type I interferon expression in blood has also been associated with anti-TNF response in RA patients, there may be an opportunity to consider the shared biology of anti-TNF response in IBD and RA. Much work has been done generating and validating signatures for anti-TNF response in RA [412]; but not much work on validating RA signatures in IBD cohorts and vice versa.

This chapter has been purely descriptive. Although there are expression differences at many genes between responders and non-responders, I do not know which cause non-response, and which are a consequence of disease reduction in responders. I have deliberately avoided the term “signature” in describing my own results, as I have not yet had a chance to assess the predictive capability of associated gene modules. I also did not find evidence for many strong and interpretable reQTL effects over time in whole blood, and therefore was unable to form hypotheses on the genetic mechanisms influencing anti-TNF response via expression. However, the presence of eQTLs for most genes and the presence of strong differences in expression post-induction may allow testing for causal mechanisms where genotype affects drug response via expression. Strategies for moving on to both prediction and causal inference will be discussed in Chapter 5.



Chapter 5

Discussion

Human immune response to perturbation is variable at numerous molecular and phenotypic levels. In this thesis, I profiled the transcriptomic response to *in vivo* vaccine and drug perturbations, established associations between expression and phenotypic response, and mapped changes over time in the genetic regulation of expression response. Chapter 2 focused on transcriptomic response to Pandemrix vaccination in the Human Immune Response Dynamics (HIRD) cohort, describing the transition from innate to adaptive immune response, and detecting associations between expression and antibody response. In Chapter 3, I considered the impact of host genetics on vaccine response in HIRD, identifying genetic variants associated with changes in expression post-vaccination, then exploring potential mechanisms explaining those associations. Finally, Chapter 4 applied similar analysis frameworks in a different context, response to anti-tumour necrosis factor (TNF) therapy in Crohn's disease (CD) patients in the Personalised Anti-TNF Therapy in Crohn's Disease (PANTS) cohort, finding distinct trajectories of expression between primary responders and non-responders to treatment. Each chapter presented its results and limitations in turn, but similarities in design and analysis qualify them for a joint deliberation. In this final chapter, I highlight shared themes, examine core limitations, and outline considerations for the design and analysis of future longitudinal *in vivo* perturbation studies to better our biological understanding of immune response to vaccines and drugs.

5.1 Strategies for detecting robust associations

In Chapters 2 and 4, I focused on identifying genes with differential expression after immune perturbation, or expression associated with phenotypic response variables—antibody titres and clinical anti-TNF response respectively. Vaccine and drug perturbation had strong effects on large proportions of the blood immune transcriptome, resulting in thousands of highly significant associations when comparing pre- and post-perturbation timepoints. In comparison, it was much more challenging to identify robust single-gene associations with response phenotypes. In Chapter 2, associations of day 7 expression with antibody response from Sobolev *et al.* [162] were replicated in my analysis of the original array data, but not in newly-generated RNA sequencing (RNA-seq) data, or in the meta-analysis. In Chapter 4, baseline associations with anti-TNF response from the literature—including at *TREM1*, previously reported by two independent groups [366, 367]—were not significant in my analysis of the PANTS cohort. The biological

effect size of a single gene's expression on phenotypic response is likely to be small, eclipsed by other sources of variation: measurement platform, difference in response definitions, sample characteristics, and noise. The idealistic suggestion is to increase sample size, but resource and ethical constraints will always exist. Rather than creating new cohorts, a logically-efficient strategy is sampling from individuals enrolled in drug and vaccine trials, but care must be taken to ensure the trial is powered both for its primary endpoints, and for planned transcriptomic analyses. Power calculations for differential gene expression (DGE) are non-trivial and it is often unknown what a reasonable effect size to assume might be. Many experiments choose parameters like sample size and sequencing depth based on rules of thumb [196], or to be comparable to existing ones in the field. In cases where small effects are likely and high power is not guaranteed, one should be cognizant of winner's curse when reporting and interpreting associations determined to be significant based on some threshold [305].

Another consideration is how best to distribute a fixed sample size between depth (number of individuals) and richness (number of timepoints, phenotypes, data types). Some degree of longitudinal sampling is recommended for a phenotype as dynamic as immune response. Chapter 2 demonstrated a distinct jump from day 1 innate to day 7 adaptive immune expression profiles post-vaccination, but the kinetics of the transition are not clear. In hindsight, responses could have peaked earlier or later in different individuals, and variation in the speed of response cannot be examined without denser sampling. In Chapter 4, expression differences between anti-TNF responders and non-responders were apparent from week 14, but it is not known if differences actually appear much earlier. Future analysis of a (small) number of available PANTS RNA-seq samples from day 3 after initiating treatment may uncover associations in the early innate response.

Rich sampling also offers analytical advantages. Having repeated measures from the same individuals allowed modelling of within-individual covariance in Chapters 2 and 4, improving statistical efficiency. The spline model in Chapter 4 enabled separation of responders and non-responders based on expression trajectory over multiple timepoints. However, all those models only incorporated two data types: expression and phenotypic response. Studies in the systems vaccinology field have demonstrated how integrating networks of many data types identifies correlates and predictors of response not only in the transcriptome, but in multiple layers of the immune system [413]. In HIRD, longitudinal fluorescence-activated cell sorting (FACS) and cytokine measurements are available for this form of integrative modelling.

When transcription is quantified on a global scale, analyses should not consider genes in isolation. Genes in the immune system are not independent, and just as variation increases uncertainty, covariation reduces it*. In Chapter 2, imprecise estimates from multiple genes were used to build an informative empirical prior for between-platform heterogeneity. Throughout the thesis, I make extensive use of enrichment analyses with gene sets defined by prior biological knowledge, to detect subtle but coordinated changes based on the expression of multiple genes. General purpose gene sets may be less relevant in immune cells [157], so I used blood transcription modules (BTMs) [239, 240] tailored for immune gene expression in blood. Alternative databases

*Wickham, H. & Grolemund, G. Chapter 7: Exploratory Data Analysis. R for Data Science. <https://r4ds.had.co.nz/exploratory-data-analysis.html>

that provide immune-focused gene sets include InnateDB [414] and MSigDB [415]. Many significant module associations with vaccine antibody response and clinical drug response were identified in Chapters 2 and 4, and my expectation is that these should be more replicable than any single-gene associations I reported (e.g. *SIGLEC10* from Chapter 4). While the effect size of a single gene may vary from sample to sample due to noise, a summary measure computed from multiple genes should be more robust. Indeed, some module associations between baseline expression and antibody response found in Chapter 2 were reported in previous studies of seasonal influenza vaccines. Most systems vaccinology studies aiming to identify consistent associations with vaccine response over multiple cohorts and sampling years focus their analyses at the gene set level [146]. Gene set analyses cannot, however, be divorced from examining the genes within them, as the genes that drive set-level associations can differ between apparent replications, and the mapping between genes and gene sets is one-to-many.

5.2 Responder analysis

A key determinant of how well the models in this thesis might correspond to reality lies in the assumed model for phenotypic response. By encoding response as an independent variable, an assumption is made that it is a stable characteristic of an individual that is measured without error*. This may not be an accurate assumption. Imagine a hypothetical drug or vaccine where 60 % of a sample of individuals have an observed response phenotype: “60 % of the time, it works every time”†. This is compatible with a stable 60 % success rate in 100 % of individuals (variation in observed response is entirely due to chance), or a stable 100 % success rate in 60 % of individuals and a 0 % success rate in the other 40 % of individuals (response is highly personal)—most likely the truth is somewhere in between. In the first scenario, it is difficult to imagine identifying robust baseline associations with response. This has been extensively discussed in the context of randomised controlled trials [416], but similar issues pertain to response definitions in observational studies.

One needs to establish how correlated phenotypic response is over time within the same individual, and computing within-individual variation requires replication at the level of the individual. The same individual must be *perturbed and measured* more than once [417]. This is not always possible in practice; in Chapter 2, antibody response was defined based on a single measurement after a single vaccine dose, but measuring response after a hypothetical second dose would quantify a different phenotype: the secondary immune response based on vaccine-induced immune memory. In Chapter 4, patients did receive repeated anti-TNF doses interspersed with sampling timepoints, and the expression differences between clinical responders and non-responders seen at week 14—the timepoint where clinical response was assessed—were maintained at week 30 and week 54. This suggests the initial designation of non-responders is not entirely due to chance, but due to some characteristic of patient disease state.

Even if response is actually a stable personal characteristic, one still needs to select an appropriate mathematical definition. As discussed in Section 2.2.2, a binary definition of response

*Note that the regression framework can accommodate measurement error in the context of errors-in-variables models.

†Apatow, J., McKay, A. & Ferrell, W. Anchorman: The Legend of Ron Burgundy (2004).

based on dichotomisation is inefficient and biologically implausible. I instead used the **titre response index (TRI)** in Chapters 2 and 3, a continuous change score combining **haemagglutination inhibition (HAI)** and **microneutralisation (MN)** titres, residualised on the baseline titres. In Chapter 4, the binary clinical response phenotype is based on a complex decision tree with many inputs. Defining dichotomies based on multiple inputs can lead to discontinuities and non-monotonicity in response probabilities under small changes in inputs [173]. Pragmatism did come into play when choosing these definitions. For **DGE**, the most widespread models have expression as the sole dependent variable, and encode phenotypic response variables as independent variables. Both **TRI** and the **PANTS** clinical response definition provided that single independent variable. In hindsight, variation in response definitions likely contributes to difficulties in replicating associations between studies, so it may be more sensible to model on the component phenotypes themselves (e.g. log **HAI** and **MN** titres, **C-reactive protein (CRP)** levels and **Harvey Bradshaw index (HBI)** scores).

5.3 Challenges in the interpretation of bulk expression data

Bulk expression data is a mixture of cell types with heterogeneous expression profiles. One of the largest sources of variation in bulk blood expression data is variation in immune cell composition, generated both from true variation in composition and sampling effects. The more cell type-specific a gene's expression, the more its measurement in bulk is affected by cell composition [418]. Highly cell type-specific genes can be treated as marker genes, used in deconvolution methods to estimate cell proportions in bulk samples when they are not directly measured. In Chapter 3, **xCell** [298]—while not technically a deconvolution method—was used to estimate cell type enrichment scores from array and **RNA-seq** expression data. In Chapter 4, estimates of cell proportions were computed by deconvolution of matched genome-wide methylation data. When fit as covariates in linear regression, cell abundance estimates act as precision variables for sampling noise, but additionally as mediators of the perturbation's effect on expression. In Chapter 4, I chose to run two sets of models with and without including estimates of five major immune cell proportions, gaining some information on which effects are likely driven by cell abundance, and which are driven by per-cell up or downregulation of transcription.

Using major cell populations for correction misses the contribution of rare populations [389]. For *cis*-expression quantitative trait locus (eQTL) mapping in Chapters 3 and 4, where the main concern was maximising the number of eQTLs detected, hidden factors from PEER [182] were included into models in addition to cell abundance estimates from deconvolution. PEER factors were correlated with deconvoluted cell abundances, so it is likely they capture additional variation from rarer cell types. If having interpretable covariates for cell abundance is unimportant, methods like surrogate variable analysis [419, 420] can be used to adjust for cell composition and other unmeasured technical sources of variation in **DGE** analyses also.

Interpretable covariates for cell abundances are important for considering response expression quantitative trait locus (reQTL) effects in bulk data. As discussed in Section 3.2.10, it is model misspecification to omit genotype-cell abundance interactions if the effect of genotype changes depending on cell abundance. It is even popular to use such interaction terms between genotype

and cell abundance (or a proxy of cell abundance) to discover cell type-specific eQTLs [71, 74]. *In vivo*, cell abundances are causally affected by the perturbation due to active recruitment, differentiation, and proliferation of immune cells. Consider the case where vaccine perturbation causes active proliferation of a rare cell type that is near absent at baseline, but forms a greatly increased proportion of the bulk mixture after perturbation. Any baseline eQTL specific to this cell type will appear as a reQTL at the post-perturbation timepoint, because expression of that cell type contributes more to the bulk mixture. If the eGene is not cell type-specific in its expression, adjusting for abundance of the cell type will only offset the regression lines at each timepoint, but not change their slopes relative to one another. The eGene also does not have to be differentially expressed on average, as the effect of interest is not the pre-post difference in expression, but the effect size of genotype on that difference. In Chapter 3, I found that an increase in naive classical monocytes at day 1 revealing a non-stimulus-specific but monocyte-specific eQTL, for the non-monocyte-specific gene *ADCY3*, was a plausible mechanism underlying the strongest day 1 reQTL.

An aim of the *in vivo* reQTL design is to find host genetic variants with a causal effect on response to perturbation. The crux of the issue is whether such an interpretation is justifiable: whether a difference in group-level eQTL regression slopes between baseline and post-perturbation necessarily entails a causal effect of genotype on change in expression from baseline to post-perturbation at the individual level. For the specific case of the *ADCY3* day 1 reQTL, I believe so. As the variant is an eQTL in monocytes, individuals with more copies of the effect allele have higher *ADCY3* expression per-monocyte on average. If you were to change the genotype of an individual from homozygous non-effect to homozygous effect, you *would* change their post-vaccination increase of *ADCY3*, because the exact same increase in monocyte abundance from baseline to day 1 would provide more *ADCY3* transcripts. It is less clear in the general case, as there are many possible mechanisms for an observed reQTL: a gene with an eQTL not expressed at baseline becoming detectable (power), a cell type with a cell type-specific eQTL increasing in proportion (recruitment or proliferation), the effect of a cell type-specific eQTL increasing within that cell type (activation, the canonical scenario assumed for *in vitro* stimulation), a genotype-dependent increase in cell abundance creating a reQTL for a gene with cell type-specific expression, *et cetera*. Not all of these can be ruled out just by including cell abundances as covariates in the eQTL model. Even if a large number of reQTLs can be detected by statistical interaction, as in Chapter 3, the challenge is distinguishing between these mechanistic scenarios and forming causal hypotheses. It is also unclear whether *in vivo* reQTLs provide additional utility over *in vitro* reQTLs for gene prioritisation at genome-wide association study (GWAS) loci. Theoretically, there may be effects unobservable without *in vivo* interactions in the immune system, but a systematic comparison of reQTLs detected with *in vivo* and *in vitro* stimulation has not been performed. *In vivo* reQTL studies are certainly not ineffectual at their stated goals, but cell composition does add considerable complexity to their interpretation. Although insights into the biological mechanism of the stimulation response is easier to gain when cell type abundance is controlled *in vitro*, one basic utility of *in vivo* stimulation is allowing the detection of additional cell type-specific eQTL effects in bulk data using genotype-cell type abundance interaction terms, a methodology already well-established in non-stimulated bulk

samples (e.g. [71]).

To truly control for cell composition, the best option is to control it at the study design stage. Adjusting for cell abundance in regression only attempts to estimate the effect of other predictors if cell abundance were held constant. It does not change the cell abundances to be equal—it is a change of viewpoint, not a change of data. Adjusting for abundance also cannot distinguish cell types with correlated abundance estimates. Single-cell RNA sequencing (scRNA-seq) after *in vivo* perturbation would quantify per-cell expression and cell abundance simultaneously. The technology is emerging as an alternative to bulk sequencing of FACS-sorted cells, having comparable cost, and the additional advantage of not requiring pre-defined marker sets [421]. There is flexibility in choosing to conduct DGE analysis and eQTL mapping within each cell type cluster, or to pool clusters to mimic bulk data. Paired designs that leverage the power of bulk reQTL mapping and the cell type resolution of single-cell data have been explored, using eGene expression in clusters to annotate bulk reQTLs to likely cell types [90]. As an emerging technology, scRNA-seq still faces many limitations, such as low coverage of the transcriptome due to dropout, smaller sample sizes due to cost, difficulties in defining robust cell type clusters, and sample processing effects on the transcriptome, but progress in the field has been nothing but rapid.

5.4 From association to prediction

In the DGE regression models I used to test for association of expression with phenotypic response, expression was always placed as the dependent variable, and response as an independent variable. In a clinical setting, a more relevant concern is prediction of patient response from expression (ideally baseline expression), reversing the roles of expression and response in the model. In Chapters 2 and 4, I observed few significant single-gene associations with response at baseline. It is first useful to consider what implications this has on the move from association to prediction in these datasets.

Prediction from genome-wide transcriptomic data is often a $p \gg n$ prediction problem, where the number of potential predictors p dwarfs the sample size n . Efron [422] provides a fascinating case study on predicting prostate cancer status from expression array data ($p = 6033$ genes) in samples from 52 prostate cancer patients and 50 controls ($n = 102$). After randomly splitting the data into training and test sets, each with 26 cancer patients and 25 controls, a random forest used to predict cancer status from gene expression recorded a 2% test set error. Repeating over many random splits showed this high predictive performance was not an outlier. Random forests have embedded feature selection, assigning their predictors an importance score, with a positive importance score indicating that a predictor was utilised by the model. After removing all 348 genes with positive importance scores in the trained model from the dataset, then repeating the process with remaining $p = 5685$ by $n = 102$ matrix, another model was produced where a set of 364 genes with positive importance—completely disjoint from the first 348—predicted cancer status with a similar error rate. This process could be repeated multiple times, each time producing a model with similar error rate, using none of the “important” genes from the previous models. Although these error rates come from internal validation, which have an optimistic

bias, the performance of pure prediction models does appear to be dominated by the confluence of many weak predictors. Therefore it is still feasible to consider prediction in datasets where attribution of significance to individual strong predictors may be impossible.

A large part of systems vaccinology in the last decade has been building models to predict vaccine-induced antibody and cellular responses from high-dimensional data. The methods used span the full gamut of statistical and machine learning algorithms, including classification to nearest centroid (ClaNC) [249], discriminant analysis via mixed integer programming (DAMIP) [154, 157, 249, 423], nearest shrunken centroid algorithm (e.g. PAM [149]), linear regression [153, 413], logistic regression [127, 155, 424], linear discriminant analysis (LDA) [161, 425], elastic net [160], partial least squares (PLS) [161], artificial neural networks (ANN) [157], naive Bayes [426], lasso regression [427], sparse partial least squares (SPLS) [428], and logistic multiple network-constrained regression (LogMiNeR) [429, 430]. The choice of methodology can be daunting. Fortunately (or unfortunately), an extensive survey of transcriptomic prediction models by the MicroArray Quality Control Consortium [431] found that the choice of algorithm was not as influential on predictive performance as the endpoint itself, with some endpoints being inherently difficult to predict. There is also no need to restrict oneself to a particular method; ensemble models that combine multiple algorithms consistently have the best performance and robustness [432]. It is hard to say *a priori* whether antibody response in HIRD and anti-TNF response in PANTS are “difficult” endpoints. The existence of predictive signatures for seasonal influenza vaccine response using baseline expression—validated over multiple cohorts, years, and geographical locations—does set an encouraging precedent for the former [159].

Oncology was one of the earliest fields to adopt predictive gene signatures into clinical practice. Despite the first commercial tests launching in the early 2000s (e.g. MammaPrint, a 76-gene signature for breast cancer prognosis), only a handful are in use today [433–435]. There are multiple hurdles to clinical implementation, requiring that a signature not only have validated accuracy, but provide sufficient incremental value on top of existing clinical markers in a cost-effective manner [434]. Feature selection is of particular importance when building models for the clinic; cost-effectiveness entails that most expression tests are qPCR-based tests that measure at most a few dozen genes. There is an interesting tension between the sparsity assumed by feature selection methods (that most predictors have no effect) and the observation that prediction algorithms depend on many weak predictors. A balance between predictive performance and measurement cost will likely need to be struck. The ability to predict individual response to anti-TNF treatment would be revolutionary due to the treatment cost and quality of life impact of taking ineffectual biologic therapy. The case for personalised vaccinology lies mostly in building understanding of the best vaccine type, dose, and timing for vaccination of challenging populations [124].

5.5 From association to causality

Knowing the causal mechanisms of immune response to perturbation is crucial for conceiving of possible interventions. For example, assuming the baseline association of *SIGLEC10* with anti-TNF response identified in Chapter 4 is a true association, would intervening on baseline

SIGLEC10 expression affect probability of response? The study designs used in this thesis are uncontrolled, but still provide useful guarantees against reverse causality. Post-perturbation phenotypic or expression measurements cannot cause baseline gene expression. Post-conception phenotypic or expression measurements cannot cause genotype. To estimate causal effects of expression on phenotype, models are needed that encode causal relationships as testable hypotheses. There are several families of such methods; as I shall now describe, they should be used in combination.

Mendelian randomisation (MR) is a form of instrumental variable (IV) analysis that uses genetic variants as IVs to estimate the causal effect of an exposure on an outcome. Three assumptions define a valid genetic IV [436–438]. In the case where the exposure is gene expression, and the outcome is some phenotypic response such as antibody titre, the first assumption (IV1) is that the variant should be associated with the exposure as an eQTL. The term MR comes from an analogy to randomised controlled trials; meiotic segregation is largely independent of environmental confounders, so different eQTL alleles can be thought to randomly assign different “doses” of expression [436]. The second assumption (IV2) is that the variant is not associated with unmeasured confounders of the expression-phenotype association (e.g. population structure). The third assumption (IV3) is that the variant has no association with phenotype except through expression. Combined, these assumptions place expression as a complete mediator (vertical pleiotropy) of the effect of the eQTL on phenotype (Fig. 5.1). The effects of variant on expression and expression on phenotype can be estimated in the same sample, or in non-overlapping samples (two-sample MR [437, 438]). Two-sample MR can leverage existing large eQTL catalogues and helps mitigate weak-instrument bias, where eQTLs with weak effects on expression are used as IVs. The direction of bias in estimating the expression-phenotype effect is away from the null in single-sample MR, but towards the null in two-sample MR [439, 440]. A related family of methods, transcriptome-wide association studies (TWASs) [441], train predictive models of expression from eQTL data, then apply those models in GWAS cohorts to test the association of genetically-predicted expression with phenotype. TWAS methods have methodological similarities to two-sample MR [442].

Violating the assumptions of MR will likely lead to biased causal estimates. The most troublesome assumption is often IV3. If there is no temporal ordering of exposure and outcome, IV3 can be violated by reverse causation. For example, if evaluating the causal effect of day 7 post-vaccination gene expression on day 7 CD4⁺ T cell abundance, an association between variant and expression might be mediated by the phenotype, cell abundance. If this is suspected, one can perform MR in the reverse direction if there are available instruments for the phenotype (bi-directional MR), or perform a statistical test of the directionality (MR Steiger) [436–438, 443]. IV3 can be violated by linkage if the eQTL does not actually have any effect on the phenotype at all, but simply is in linkage disequilibrium (LD) with another variant that does; and can also be violated by the existence of horizontal pleiotropy, where the effect of the variant on expression and phenotype are independent (Fig. 5.1). Colocalisation methods, as used in Chapter 3, can be used to test whether the same causal variant affects expression and phenotype, distinguishing pleiotropy from linkage. However, colocalisation is necessary but not sufficient for mediation, thus it does not distinguish mediation (vertical pleiotropy) from horizontal pleiotropy [437].

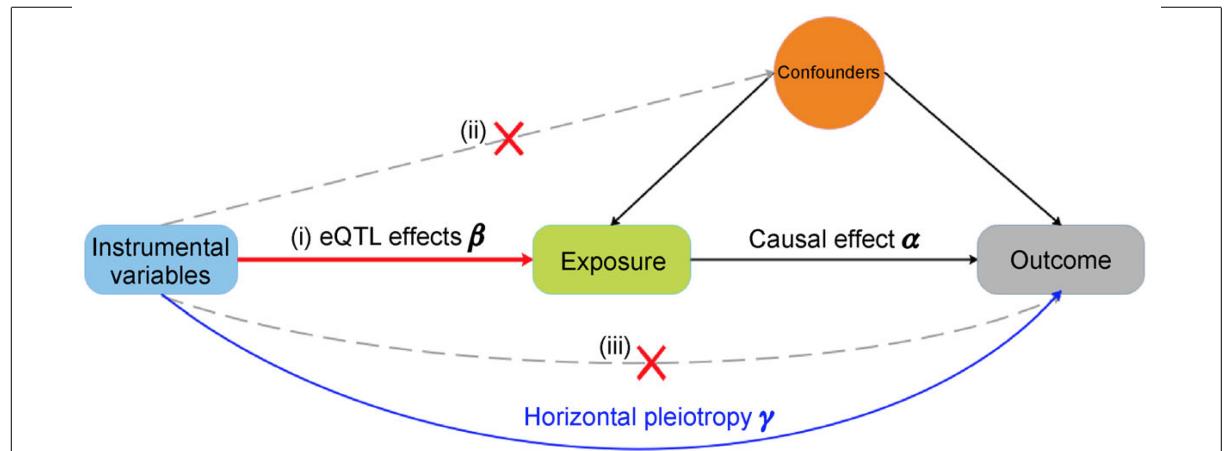


Figure 5.1: The three assumptions of MR. MR uses genetic IVs to estimate the causal effect α of an exposure (here, gene expression) on a phenotypic outcome, under three assumptions: (i) IV1: the variant is associated with the exposure (here, an eQTL with effect size β); (ii) IV2: the variant is not associated with any unmeasured confounders; (iii) IV3: the variant is not associated with the outcome except through exposure. The directionality of the arrows in the causal diagram are also assumed to hold. The blue arrow shows a horizontal pleiotropic effect of the variant on outcome, a violation of the IV3 assumption. Figure reprinted by permission from Springer Nature: Springer Nature, Quantitative Biology, Zhu *et al.* [442], © 2020.

Mediation analysis methods (e.g. CIT [263], Findr [444]) can be used to test for violations of IV3 by horizontal pleiotropy. They distinguish mediation from horizontal pleiotropy using comparison of causal models with different structures, but require individual level data, and are more susceptible to measurement error than MR [437, 443].

5.6 Triangulation

Triangulation refers to the use of methods that address the same question, but with different assumptions, biases, and limitations [445]. An example from this thesis appears in Chapter 3, combining DGE, between-individual reQTL mapping, and colocalisation—and pending validation by within-individual allele-specific expression (ASE)—to propose mechanisms behind changes in the genetic architecture of immune gene expression after vaccination. As discussed above, MR, colocalisation, and mediation analysis can be seen as complementary methods for triangulating the causal relationships between variant, exposure, and outcome. Taylor *et al.* [446] and Zheng *et al.* [447] exemplify how these methods can be combined in practice for genetic instruments, molecular exposures, and molecular outcomes. A combination of methods addresses limitations that cannot be solved by increasing sample size. Triangulation will be critical in moving from a descriptive to a mechanistic understanding of immune response to perturbations.

5.7 Concluding remarks

It has now been almost two decades since the completion of the Human Genome Project and the conception of systems biology, and almost fifteen years since the first GWASs and systems immunology studies. High-throughput profiling, complex algorithms, and big data are the new normal, yet the classical principles of perturbation and observation are alive and well. The

projects in this thesis come in the wake of these monumental achievements, yet still lie at the beginning of a long road leading towards a fuller understanding of our immune system.

The goal must be to not only observe the immune response to perturbation, but to be able to predict it, and to understand the causal relationships within the immune system that will ultimately guide the rational design and administration of vaccines and drugs. For this, we need study designs and analysis strategies for detecting robust and replicable associations with sensible response phenotypes. We need technologies that quantify the immune system with great richness and resolution, yet remain affordable enough to do so without sacrificing sample size. We need triangulation via multiple lines of evidence, requiring both confluence of methodology and collaboration of minds. The road from perturbation to understanding is a long one indeed, but it shall be a road paved by good science.

Bibliography

1. Polderman, T. J. C. *et al.* Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies. *Nature Genetics* **47**, 702–709. doi:[10.1038/ng.3285](https://doi.org/10.1038/ng.3285) (2015).
2. Lawson, D. J. *et al.* Is Population Structure in the Genetic Biobank Era Irrelevant, a Challenge, or an Opportunity? *Human Genetics* **139**, 23–41. doi:[10.1007/s00439-019-02014-8](https://doi.org/10.1007/s00439-019-02014-8) (2020).
3. Day, F. R., Loh, P.-R., Scott, R. A., Ong, K. K. & Perry, J. R. A Robust Example of Collider Bias in a Genetic Association Study. *The American Journal of Human Genetics* **98**, 392–393. doi:[10.1016/j.ajhg.2015.12.019](https://doi.org/10.1016/j.ajhg.2015.12.019) (2016).
4. The ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **489**, 57–74. doi:[10.1038/nature11247](https://doi.org/10.1038/nature11247) (2012).
5. 1000 Genomes Project Consortium *et al.* A Global Reference for Human Genetic Variation. *Nature* **526**, 68–74. doi:[10.1038/nature15393](https://doi.org/10.1038/nature15393) (2015).
6. The International SNP Map Working Group. A Map of Human Genome Sequence Variation Containing 1.42 Million Single Nucleotide Polymorphisms. *Nature* **409**, 928–933. doi:[10.1038/35057149](https://doi.org/10.1038/35057149) (2001).
7. Slatkin, M. Linkage Disequilibrium — Understanding the Evolutionary Past and Mapping the Medical Future. *Nature Reviews Genetics* **9**, 477–485. doi:[10.1038/nrg2361](https://doi.org/10.1038/nrg2361) (2008).
8. Wall, J. D. & Pritchard, J. K. Haplotype Blocks and Linkage Disequilibrium in the Human Genome. *Nature Reviews Genetics* **4**, 587–597. doi:[10.1038/nrg1123](https://doi.org/10.1038/nrg1123) (2003).
9. The International HapMap Consortium. A Second Generation Human Haplotype Map of over 3.1 Million SNPs. *Nature* **449**, 851–861. doi:[10.1038/nature06258](https://doi.org/10.1038/nature06258) (2007).
10. Karczewski, K. J. & Martin, A. R. Analytic and Translational Genetics. *Annual Review of Biomedical Data Science* **3**. doi:[10.1146/annurev-biodatasci-072018-021148](https://doi.org/10.1146/annurev-biodatasci-072018-021148) (2020).
11. Pääbo, S. The Mosaic That Is Our Genome. *Nature* **421**, 409–412. doi:[10.1038/nature01400](https://doi.org/10.1038/nature01400) (2003).
12. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic Architecture: The Shape of the Genetic Contribution to Human Traits and Disease. *Nature Reviews Genetics* **19**, 110–124. doi:[10.1038/nrg.2017.101](https://doi.org/10.1038/nrg.2017.101) (2018).
13. Visscher, P. M. & Goddard, M. E. From R.A. Fisher's 1918 Paper to GWAS a Century Later. *Genetics* **211**, 1125–1130. doi:[10.1534/genetics.118.301594](https://doi.org/10.1534/genetics.118.301594) (2019).

14. Hindorff, L. A. *et al.* Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits. *Proceedings of the National Academy of Sciences* **106**, 9362–9367. doi:[10.1073/pnas.0903103106](https://doi.org/10.1073/pnas.0903103106) (2009).
15. Gibson, G. Rare and Common Variants: Twenty Arguments. *Nature reviews. Genetics* **13**, 135–145. doi:[10.1038/nrg3118](https://doi.org/10.1038/nrg3118) (2011).
16. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186. doi:[10.1016/j.cell.2017.05.038](https://doi.org/10.1016/j.cell.2017.05.038) (2017).
17. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic Mapping in Human Disease. *Science* **322**, 881–8. doi:[10.1126/science.1156409](https://doi.org/10.1126/science.1156409) (2008).
18. Ott, J., Kamatani, Y. & Lathrop, M. Family-Based Designs for Genome-Wide Association Studies. *Nature Reviews Genetics* **12**, 465–474. doi:[10.1038/nrg2989](https://doi.org/10.1038/nrg2989) (2011).
19. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five Years of GWAS Discovery. *The American Journal of Human Genetics* **90**, 7–24. doi:[10.1016/j.ajhg.2011.11.029](https://doi.org/10.1016/j.ajhg.2011.11.029) (2012).
20. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A Comprehensive Review of Genetic Association Studies. *Genetics in Medicine* **4**, 45–61. doi:[10.1097/GIM.0000000000000002](https://doi.org/10.1097/GIM.0000000000000002) (2002).
21. Border, R. *et al.* No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *American Journal of Psychiatry* **176**, 376–387. doi:[10.1176/appi.ajp.2018.18070881](https://doi.org/10.1176/appi.ajp.2018.18070881) (2019).
22. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* **101**, 5–22. doi:[10.1016/j.ajhg.2017.06.005](https://doi.org/10.1016/j.ajhg.2017.06.005) (2017).
23. Sinnott-Armstrong, N., Naqvi, S., Rivas, M. & Pritchard, J. K. GWAS of Three Molecular Traits Highlights Core Genes and Pathways alongside a Highly Polygenic Background. *bioRxiv*. doi:[10.1101/2020.04.20.051631](https://doi.org/10.1101/2020.04.20.051631) (2020).
24. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G. & Meyre, D. Benefits and Limitations of Genome-Wide Association Studies. *Nature Reviews Genetics*. doi:[10.1038/s41576-019-0127-1](https://doi.org/10.1038/s41576-019-0127-1) (2019).
25. Crouch, D. J. M. & Bodmer, W. F. Polygenic Inheritance, GWAS, Polygenic Risk Scores, and the Search for Functional Variants. *Proceedings of the National Academy of Sciences* **117**, 18924–18933. doi:[10.1073/pnas.2005634117](https://doi.org/10.1073/pnas.2005634117) (2020).
26. Bush, W. S. & Moore, J. H. Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology* **8** (eds Lewitter, F. & Kann, M.) e1002822. doi:[10.1371/journal.pcbi.1002822](https://doi.org/10.1371/journal.pcbi.1002822) (2012).
27. The International HapMap Consortium. A Haplotype Map of the Human Genome. *Nature* **437**, 1299–1320. doi:[10.1038/nature04226](https://doi.org/10.1038/nature04226) (2005).
28. Barrett, J. C. & Cardon, L. R. Evaluating Coverage of Genome-Wide Association Studies. *Nature Genetics* **38**, 659–662. doi:[10.1038/ng1801](https://doi.org/10.1038/ng1801) (2006).

29. Das, S., Abecasis, G. R. & Browning, B. L. Genotype Imputation from Large Reference Panels. *Annual Review of Genomics and Human Genetics* **19**, 73–96. doi:[10.1146/annurev-genom-083117-021602](https://doi.org/10.1146/annurev-genom-083117-021602) (2018).
30. Taliun, D. *et al.* Sequencing of 53,831 Diverse Genomes from the NHLBI TOPMed Program. *bioRxiv*. doi:[10.1101/563866](https://doi.org/10.1101/563866) (2019).
31. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the Multiple Testing Burden for Genomewide Association Studies of Nearly All Common Variants. *Genetic Epidemiology* **32**, 381–385. doi:[10.1002/gepi.20303](https://doi.org/10.1002/gepi.20303) (2008).
32. Jannet, A.-S., Ehret, G. & Perneger, T. $P < 5 \times 10^{-8}$ Has Emerged as a Standard of Statistical Significance for Genome-Wide Association Studies. *Journal of Clinical Epidemiology* **68**, 460–465. doi:[10.1016/j.jclinepi.2015.01.001](https://doi.org/10.1016/j.jclinepi.2015.01.001) (2015).
33. Goeman, J. J. & Solari, A. Multiple Hypothesis Testing in Genomics. *Statistics in Medicine* **33**, 1946–1978. doi:[10.1002/sim.6082](https://doi.org/10.1002/sim.6082) (2014).
34. Schaid, D. J., Chen, W. & Larson, N. B. From Genome-Wide Associations to Candidate Causal Variants by Statistical Fine-Mapping. *Nature Reviews Genetics* **19**, 491–504. doi:[10.1038/s41576-018-0016-z](https://doi.org/10.1038/s41576-018-0016-z) (2018).
35. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. doi:[10.1111/rssb.12388](https://doi.org/10.1111/rssb.12388) (2020).
36. Chong, J. X. *et al.* The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *The American Journal of Human Genetics* **97**, 199–215. doi:[10.1016/j.ajhg.2015.06.009](https://doi.org/10.1016/j.ajhg.2015.06.009) (2015).
37. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *The American Journal of Human Genetics* **102**, 717–730. doi:[10.1016/j.ajhg.2018.04.002](https://doi.org/10.1016/j.ajhg.2018.04.002) (2018).
38. Brodie, A., Azaria, J. R. & Ofran, Y. How Far from the SNP May the Causative Genes Be? *Nucleic Acids Research* **44**, 6046–6054. doi:[10.1093/nar/gkw500](https://doi.org/10.1093/nar/gkw500) (2016).
39. Stacey, D. *et al.* ProGeM: A Framework for the Prioritization of Candidate Causal Genes at Molecular Quantitative Trait Loci. *Nucleic Acids Research* **47**, e3–e3. doi:[10.1093/nar/gky837](https://doi.org/10.1093/nar/gky837) (2019).
40. Forgetta, V. *et al.* An Effector Index to Predict Causal Genes at GWAS Loci. *bioRxiv*. doi:[10.1101/2020.06.28.171561](https://doi.org/10.1101/2020.06.28.171561) (2020).
41. Ghoussaini, M. *et al.* Open Targets Genetics: Systematic Identification of Trait-Associated Genes Using Large-Scale Genetics and Functional Genomics. *Nucleic Acids Research*. doi:[10.1093/nar/gkaa840](https://doi.org/10.1093/nar/gkaa840) (2020).
42. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics* **11**. doi:[10.3389/fgene.2020.00424](https://doi.org/10.3389/fgene.2020.00424) (2020).

43. Cramer, P. Organization and Regulation of Gene Transcription. *Nature* **573**, 45–54. doi:[10.1038/s41586-019-1517-4](https://doi.org/10.1038/s41586-019-1517-4) (2019).
44. Bannister, A. J. & Kouzarides, T. Regulation of Chromatin by Histone Modifications. *Cell Research* **21**, 381–395. doi:[10.1038/cr.2011.22](https://doi.org/10.1038/cr.2011.22) (2011).
45. Robertson, K. D. & Wolffe, A. P. DNA Methylation in Health and Disease. *Nature Reviews Genetics* **1**, 11–19. doi:[10.1038/35049533](https://doi.org/10.1038/35049533) (2000).
46. Gott, J. M. & Emeson, R. B. Functions and Mechanisms of RNA Editing. *Annual Review of Genetics* **34**, 499–531. doi:[10.1146/annurev.genet.34.1.499](https://doi.org/10.1146/annurev.genet.34.1.499) (2000).
47. Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated Regions of mRNAs. *Genome Biology* **3**, REVIEWS0004. doi:[10.1186/gb-2002-3-3-reviews0004](https://doi.org/10.1186/gb-2002-3-3-reviews0004) (2002).
48. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking Disease Associations with Regulatory Information in the Human Genome. *Genome Research* **22**, 1748–1759. doi:[10.1101/gr.136127.111](https://doi.org/10.1101/gr.136127.111) (2012).
49. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195. doi:[10.1126/science.1222794](https://doi.org/10.1126/science.1222794) (2012).
50. Farh, K. K.-H. *et al.* Genetic and Epigenetic Fine Mapping of Causal Autoimmune Disease Variants. *Nature* **518**, 337–343. doi:[10.1038/nature13835](https://doi.org/10.1038/nature13835) (2015).
51. Trynka, G. *et al.* Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-Coding Variants within Complex-Trait Loci. *The American Journal of Human Genetics* **97**, 139–152. doi:[10.1016/j.ajhg.2015.05.016](https://doi.org/10.1016/j.ajhg.2015.05.016) (2015).
52. Nasser, J. *et al.* Genome-Wide Maps of Enhancer Regulation Connect Risk Variants to Disease Genes. *bioRxiv*. doi:[10.1101/2020.09.01.278093](https://doi.org/10.1101/2020.09.01.278093) (2020).
53. Gaffney, D. J. Global Properties and Functional Complexity of Human Gene Regulatory Variation. *PLoS Genetics* **9** (ed Abecasis, G. R.) e1003501. doi:[10.1371/journal.pgen.1003501](https://doi.org/10.1371/journal.pgen.1003501) (2013).
54. The GTEx Consortium. The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. *Science* **369**, 1318–1330. doi:[10.1126/science.aaz1776](https://doi.org/10.1126/science.aaz1776) (2020).
55. Vandiedonck, C. Genetic Association of Molecular Traits: A Help to Identify Causative Variants in Complex Diseases. *Clinical Genetics*. doi:[10.1111/cge.13187](https://doi.org/10.1111/cge.13187) (2017).
56. Westra, H.-J. & Franke, L. From Genome to Function by Studying eQTLs. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1842**, 1896–1902. doi:[10.1016/j.bbadi.2014.04.024](https://doi.org/10.1016/j.bbadi.2014.04.024) (2014).
57. Albert, F. W. & Kruglyak, L. The Role of Regulatory Variation in Complex Traits and Disease. *Nature Reviews Genetics* **16**, 197–212. doi:[10.1038/nrg3891](https://doi.org/10.1038/nrg3891) (2015).
58. Võsa, U. *et al.* Unraveling the Polygenic Architecture of Complex Traits Using Blood eQTL Meta-Analysis. *bioRxiv*. doi:[10.1101/447367](https://doi.org/10.1101/447367) (2018).
59. Fairfax, B. P. *et al.* Genetics of Gene Expression in Primary Immune Cells Identifies Cell Type-Specific Master Regulators and Roles of HLA Alleles. *Nature Genetics* **44**, 502–510. doi:[10.1038/ng.2205](https://doi.org/10.1038/ng.2205) (2012).

60. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnipotent Inheritance. *Cell* **177**, 1022–1034.e6. doi:[10.1016/j.cell.2019.04.014](https://doi.org/10.1016/j.cell.2019.04.014) (2019).
61. Yao, C. *et al.* Sex- and Age-Interacting eQTLs in Human Complex Diseases. *Human Molecular Genetics* **23**, 1947–1956. doi:[10.1093/hmg/ddt582](https://doi.org/10.1093/hmg/ddt582) (2014).
62. De Jager, P. L., Hacohen, N., Mathis, D., Regev, A., Stranger, B. E. & Benoist, C. ImmVar Project: Insights and Design Considerations for Future Studies of “Healthy” Immune Variation. *Seminars in Immunology* **27**, 51–57. doi:[10.1016/j.smim.2015.03.003](https://doi.org/10.1016/j.smim.2015.03.003) (2015).
63. Nédélec, Y. *et al.* Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* **167**, 657–669.e21. doi:[10.1016/j.cell.2016.09.025](https://doi.org/10.1016/j.cell.2016.09.025) (2016).
64. Quach, H. & Quintana-Murci, L. Living in an Adaptive World: Genomic Dissection of the Genus Homo and Its Immune Response. *Journal of Experimental Medicine* **214**, 877–894. doi:[10.1084/jem.20161942](https://doi.org/10.1084/jem.20161942) (2017).
65. Nica, A. C. *et al.* The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study. *PLoS Genetics* **7** (ed Barsh, G.) e1002003. doi:[10.1371/journal.pgen.1002003](https://doi.org/10.1371/journal.pgen.1002003) (2011).
66. Aguet, F. *et al.* Genetic Effects on Gene Expression across Human Tissues. *Nature* **550**, 204–213. doi:[10.1038/nature24277](https://doi.org/10.1038/nature24277) (2017).
67. Dimas, A. S. *et al.* Common Regulatory Variation Impacts Gene Expression in a Cell Type-Dependent Manner. *Science* **325**, 1246–1250. doi:[10.1126/science.1174148](https://doi.org/10.1126/science.1174148) (2009).
68. Peters, J. E. *et al.* Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. *PLOS Genetics* **12** (ed Plagnol, V.) e1005908. doi:[10.1371/journal.pgen.1005908](https://doi.org/10.1371/journal.pgen.1005908) (2016).
69. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398–1414.e24. doi:[10.1016/j.cell.2016.10.026](https://doi.org/10.1016/j.cell.2016.10.026) (2016).
70. Calderon, D. *et al.* Landscape of Stimulation-Responsive Chromatin across Diverse Human Immune Cells. *Nature Genetics* **51**, 1494–1505. doi:[10.1038/s41588-019-0505-9](https://doi.org/10.1038/s41588-019-0505-9) (2019).
71. Westra, H.-J. *et al.* Cell Specific eQTL Analysis without Sorting Cells. *PLOS Genetics* **11** (ed Pastinen, T.) e1005223. doi:[10.1371/journal.pgen.1005223](https://doi.org/10.1371/journal.pgen.1005223) (2015).
72. Zhernakova, D. V. *et al.* Identification of Context-Dependent Expression Quantitative Trait Loci in Whole Blood. *Nature Genetics* **49**, 139–145. doi:[10.1038/ng.3737](https://doi.org/10.1038/ng.3737) (2017).
73. Glastonbury, C. A., Couto Alves, A., El-Sayed Moustafa, J. S. & Small, K. S. Cell-Type Heterogeneity in Adipose Tissue Is Associated with Complex Traits and Reveals Disease-Relevant Cell-Specific eQTLs. *The American Journal of Human Genetics* **104**, 1013–1024. doi:[10.1016/j.ajhg.2019.03.025](https://doi.org/10.1016/j.ajhg.2019.03.025) (2019).
74. Kim-Hellmuth, S. *et al.* Cell Type-Specific Genetic Regulation of Gene Expression across Human Tissues. *Science* **369**, eaaz8528. doi:[10.1126/science.aaz8528](https://doi.org/10.1126/science.aaz8528) (2020).
75. Strober, B. J. *et al.* Dynamic Genetic Regulation of Gene Expression during Cellular Differentiation. *Science* **364**, 1287–1290. doi:[10.1126/science.aaw0040](https://doi.org/10.1126/science.aaw0040) (2019).

76. Fu, J. *et al.* Unraveling the Regulatory Mechanisms Underlying Tissue-Dependent Genetic Variation of Gene Expression. *PLoS Genetics* **8** (ed Gibson, G.) e1002431. doi:[10.1371/journal.pgen.1002431](https://doi.org/10.1371/journal.pgen.1002431) (2012).
77. Huang, Q. *The Genetics of Gene Expression: From Simulations to the Early-Life Origins of Immune Diseases* (2019).
78. Barreiro, L. B., Tailleux, L., Pai, A. A., Gicquel, B., Marioni, J. C. & Gilad, Y. Deciphering the Genetic Architecture of Variation in the Immune Response to Mycobacterium Tuberculosis Infection. *Proceedings of the National Academy of Sciences* **109**, 1204–1209. doi:[10.1073/pnas.1115761109](https://doi.org/10.1073/pnas.1115761109) (2012).
79. Fairfax, B. P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science* **343**, 1246949–1246949. doi:[10.1126/science.1246949](https://doi.org/10.1126/science.1246949) (2014).
80. Kim, S. *et al.* Characterizing the Genetic Basis of Innate Immune Response in TLR4-Activated Human Monocytes. *Nature Communications* **5**. doi:[10.1038/ncomms6236](https://doi.org/10.1038/ncomms6236) (2014).
81. Hu, X. *et al.* Regulation of Gene Expression in Autoimmune Disease Loci and the Genetic Basis of Proliferation in CD4+ Effector Memory T Cells. *PLoS Genetics* **10** (ed Roopenian, D. C.) e1004404. doi:[10.1371/journal.pgen.1004404](https://doi.org/10.1371/journal.pgen.1004404) (2014).
82. Lee, M. N. *et al.* Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. *Science* **343**, 1246980–1246980. doi:[10.1126/science.1246980](https://doi.org/10.1126/science.1246980) (2014).
83. Çalışkan, M., Baker, S. W., Gilad, Y. & Ober, C. Host Genetic Variation Influences Gene Expression Response to Rhinovirus Infection. *PLOS Genetics* **11** (ed Gibson, G.) e1005111. doi:[10.1371/journal.pgen.1005111](https://doi.org/10.1371/journal.pgen.1005111) (2015).
84. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**, 643–656.e17. doi:[10.1016/j.cell.2016.09.024](https://doi.org/10.1016/j.cell.2016.09.024) (2016).
85. Kim-Hellmuth, S. *et al.* Genetic Regulatory Effects Modified by Immune Activation Contribute to Autoimmune Disease Associations. *Nature Communications* **8**. doi:[10.1038/s41467-017-00366-1](https://doi.org/10.1038/s41467-017-00366-1) (2017).
86. Alasoo, K. *et al.* Shared Genetic Effects on Chromatin and Gene Expression Indicate a Role for Enhancer Priming in Immune Response. *Nature Genetics* **50**, 424–431. doi:[10.1038/s41588-018-0046-7](https://doi.org/10.1038/s41588-018-0046-7) (2018).
87. Gate, R. E. *et al.* Genetic Determinants of Co-Accessible Chromatin Regions in Activated T Cells across Humans. *Nature Genetics* **50**, 1140–1150. doi:[10.1038/s41588-018-0156-2](https://doi.org/10.1038/s41588-018-0156-2) (2018).
88. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701–1715.e16. doi:[10.1016/j.cell.2018.10.022](https://doi.org/10.1016/j.cell.2018.10.022) (2018).

89. Alasoo, K., Rodrigues, J., Danesh, J., Freitag, D. F., Paul, D. S. & Gaffney, D. J. Genetic Effects on Promoter Usage Are Highly Context-Specific and Contribute to Complex Traits. *eLife* **8**. doi:[10.7554/eLife.41673](https://doi.org/10.7554/eLife.41673) (2019).
90. De Vries, D. H. *et al.* Integrating GWAS with Bulk and Single-Cell RNA-Sequencing Reveals a Role for LY86 in the Anti-Candida Host Response. *PLOS Pathogens* **16** (ed May, R. C.) e1008408. doi:[10.1371/journal.ppat.1008408](https://doi.org/10.1371/journal.ppat.1008408) (2020).
91. Huang, Q. Q. *et al.* Neonatal Genetics of Gene Expression Reveal Potential Origins of Autoimmune and Allergic Disease Risk. *Nature Communications* **11**. doi:[10.1038/s41467-020-17477-x](https://doi.org/10.1038/s41467-020-17477-x) (2020).
92. Manry, J. *et al.* Deciphering the Genetic Control of Gene Expression Following Mycobacterium Leprae Antigen Stimulation. *PLOS Genetics* **13** (ed Sirugo, G.) e1006952. doi:[10.1371/journal.pgen.1006952](https://doi.org/10.1371/journal.pgen.1006952) (2017).
93. Idaghdour, Y. *et al.* Evidence for Additive and Interaction Effects of Host Genotype and Infection in Malaria. *Proceedings of the National Academy of Sciences* **109**, 16786–16793. doi:[10.1073/pnas.1204945109](https://doi.org/10.1073/pnas.1204945109) (2012).
94. Franco, L. M. *et al.* Integrative Genomic Analysis of the Human Immune Response to Influenza Vaccination. *eLife* **2**, e00299. doi:[10.7554/eLife.00299](https://doi.org/10.7554/eLife.00299) (2013).
95. Lareau, C. A., White, B. C., Oberg, A. L., Kennedy, R. B., Poland, G. A. & McKinney, B. A. An Interaction Quantitative Trait Loci Tool Implicates Epistatic Functional Variants in an Apoptosis Pathway in Smallpox Vaccine eQTL Data. *Genes & Immunity* **17**, 244–250. doi:[10.1038/gene.2016.15](https://doi.org/10.1038/gene.2016.15) (2016).
96. Davenport, E. E. *et al.* Discovering in Vivo Cytokine-eQTL Interactions from a Lupus Clinical Trial. *Genome Biology* **19**. doi:[10.1186/s13059-018-1560-8](https://doi.org/10.1186/s13059-018-1560-8) (2018).
97. Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E. & Cox, N. J. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS Genetics* **6** (ed Gibson, G.) e1000888. doi:[10.1371/journal.pgen.1000888](https://doi.org/10.1371/journal.pgen.1000888) (2010).
98. Burgess, S., Foley, C. N. & Zuber, V. Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. *Annual Review of Genomics and Human Genetics* **19**, 303–327. doi:[10.1146/annurev-genom-083117-021731](https://doi.org/10.1146/annurev-genom-083117-021731) (2018).
99. Wallace, C. Eliciting Priors and Relaxing the Single Causal Variant Assumption in Colocalisation Analyses. *PLOS Genetics* **16** (ed Epstein, M. P.) e1008720. doi:[10.1371/journal.pgen.1008720](https://doi.org/10.1371/journal.pgen.1008720) (2020).
100. Hukku, A., Pividori, M., Luca, F., Pique-Regi, R., Im, H. K. & Wen, X. Probabilistic Colocalization of Genetic Variants from Complex and Molecular Traits: Promise and Limitations. *bioRxiv*. doi:[10.1101/2020.07.01.182097](https://doi.org/10.1101/2020.07.01.182097) (2020).
101. Kundu, K. *et al.* Genetic Associations at Regulatory Phenotypes Improve Fine-Mapping of Causal Variants for Twelve Immune-Mediated Diseases. doi:[10.1101/2020.01.15.907436](https://doi.org/10.1101/2020.01.15.907436) (2020).

102. Murphy, K. & Weaver, C. *Chapter 1: Basic Concepts in Immunology* in *Janeway's Immunobiology* 9th edition (Garland Science/Taylor & Francis Group, LLC, New York, NY, 2016).
103. Domínguez-Andrés, J. & Netea, M. G. The Specifics of Innate Immune Memory. *Science* **368**, 1052–1053. doi:[10.1126/science.abc2660](https://doi.org/10.1126/science.abc2660) (2020).
104. Davis, M. M., Tato, C. M. & Furman, D. Systems Immunology: Just Getting Started. *Nature Immunology* **18**, 725–732. doi:[10.1038/ni.3768](https://doi.org/10.1038/ni.3768) (2017).
105. Villani, A.-C., Sarkizova, S. & Hacohen, N. Systems Immunology: Learning the Rules of the Immune System. *Annual Review of Immunology* **36**, 813–842. doi:[10.1146/annurev-immunol-042617-053035](https://doi.org/10.1146/annurev-immunol-042617-053035) (2018).
106. Pulendran, B. & Davis, M. M. The Science and Medicine of Human Immunology, 13 (2020).
107. Yu, J., Peng, J. & Chi, H. Systems Immunology: Integrating Multi-Omics Data to Infer Regulatory Networks and Hidden Drivers of Immunity. *Current Opinion in Systems Biology* **15**, 19–29. doi:[10.1016/j.coisb.2019.03.003](https://doi.org/10.1016/j.coisb.2019.03.003) (2019).
108. Liston, A., Carr, E. J. & Linterman, M. A. Shaping Variation in the Human Immune System. *Trends in Immunology* **37**, 637–646. doi:[10.1016/j.it.2016.08.002](https://doi.org/10.1016/j.it.2016.08.002) (2016).
109. Brodin, P. & Davis, M. M. Human Immune System Variation. *Nature Reviews Immunology* **17**, 21–29. doi:[10.1038/nri.2016.125](https://doi.org/10.1038/nri.2016.125) (2017).
110. Patin, E. *et al.* Natural Variation in the Parameters of Innate Immune Cells Is Preferentially Driven by Genetic Factors. *Nature Immunology*. doi:[10.1038/s41590-018-0049-7](https://doi.org/10.1038/s41590-018-0049-7) (2018).
111. Liston, A. & Goris, A. The Origins of Diversity in Human Immunity. *Nature Immunology* **19**, 209–210. doi:[10.1038/s41590-018-0047-9](https://doi.org/10.1038/s41590-018-0047-9) (2018).
112. Tsang, J. S. Utilizing Population Variation, Vaccination, and Systems Biology to Study Human Immunology. *Trends in Immunology* **36**, 479–493. doi:[10.1016/j.it.2015.06.005](https://doi.org/10.1016/j.it.2015.06.005) (2015).
113. Yosef, N. & Regev, A. Writ Large: Genomic Dissection of the Effect of Cellular Environment on Immune Response. *Science* **354**, 64–68. doi:[10.1126/science.aaf5453](https://doi.org/10.1126/science.aaf5453) (2016).
114. Krammer, F. *et al.* Influenza. *Nature Reviews Disease Primers* **4**. doi:[10.1038/s41572-018-0002-y](https://doi.org/10.1038/s41572-018-0002-y) (2018).
115. Dhakal, S. & Klein, S. L. Host Factors Impact Vaccine Efficacy: Implications for Seasonal and Universal Influenza Vaccine Programs. *Journal of Virology* **93** (ed Coyne, C. B.) doi:[10.1128/JVI.00797-19](https://doi.org/10.1128/JVI.00797-19) (2019).
116. Krammer, F. The Human Antibody Response to Influenza A Virus Infection and Vaccination. *Nature Reviews Immunology* **19**, 383–397. doi:[10.1038/s41577-019-0143-6](https://doi.org/10.1038/s41577-019-0143-6) (2019).
117. World Health Organization. A Revision of the System of Nomenclature for Influenza Viruses: A WHO Memorandum. *Bulletin of the World Health Organization* **58**, 585–591 (1980).

118. Short, K. R., Kedzierska, K. & van de Sandt, C. E. Back to the Future: Lessons Learned From the 1918 Influenza Pandemic. *Frontiers in Cellular and Infection Microbiology* **8**. doi:[10.3389/fcimb.2018.00343](https://doi.org/10.3389/fcimb.2018.00343) (2018).
119. Garten, R. J. *et al.* Antigenic and Genetic Characteristics of Swine-Origin 2009 A(H1N1) Influenza Viruses Circulating in Humans. *Science* **325**, 197–201. doi:[10.1126/science.1176225](https://doi.org/10.1126/science.1176225) (2009).
120. World Health Organization. Standardization of terminology of the pandemic A(H1N1) 2009 virus. *Weekly Epidemiological Record* **86**, 480 (2011).
121. Houser, K. & Subbarao, K. Influenza Vaccines: Challenges and Solutions. *Cell Host & Microbe* **17**, 295–300. doi:[10.1016/j.chom.2015.02.012](https://doi.org/10.1016/j.chom.2015.02.012) (2015).
122. Bresee, J. S., Fry, A. M., Sambhara, S. & Cox, N. J. *Inactivated Influenza Vaccines* in *Plotkin's Vaccines* 456–488.e21 (Elsevier, 2018). doi:[10.1016/B978-0-323-35761-6.00031-6](https://doi.org/10.1016/B978-0-323-35761-6.00031-6).
123. Luke, C. J., Lakdawala, S. S. & Subbarao, K. *Influenza Vaccine—Live* in *Plotkin's Vaccines* 489–510.e7 (Elsevier, 2018). doi:[10.1016/B978-0-323-35761-6.00032-8](https://doi.org/10.1016/B978-0-323-35761-6.00032-8).
124. Poland, G., Ovsyannikova, I. & Kennedy, R. Personalized Vaccinology: A Review. *Vaccine* **36**, 5350–5357. doi:[10.1016/j.vaccine.2017.07.062](https://doi.org/10.1016/j.vaccine.2017.07.062) (2018).
125. Ramsay, M. *Influenza: The Green Book, Chapter 19* in *Immunisation against Infectious Disease* 1–27 (Public Health England, 2020).
126. Zimmermann, P. & Curtis, N. Factors That Influence the Immune Response to Vaccination. *Clinical Microbiology Reviews* **32**. doi:[10.1128/CMR.00084-18](https://doi.org/10.1128/CMR.00084-18) (2019).
127. Furman, D. *et al.* Systems Analysis of Sex Differences Reveals an Immunosuppressive Role for Testosterone in the Response to Influenza Vaccination. *Proceedings of the National Academy of Sciences* **111**, 869–874. doi:[10.1073/pnas.1321060111](https://doi.org/10.1073/pnas.1321060111) (2014).
128. Henry, C., Palm, A.-K. E., Krammer, F. & Wilson, P. C. From Original Antigenic Sin to the Universal Influenza Virus Vaccine. *Trends in Immunology* **39**, 70–79. doi:[10.1016/j.it.2017.08.003](https://doi.org/10.1016/j.it.2017.08.003) (2018).
129. Davis, C. W. *et al.* Influenza Vaccine-Induced Human Bone Marrow Plasma Cells Decline within a Year after Vaccination. *Science* **370**, 237–241. doi:[10.1126/science.aaz8432](https://doi.org/10.1126/science.aaz8432) (2020).
130. Sano, K., Ainai, A., Suzuki, T. & Hasegawa, H. The Road to a More Effective Influenza Vaccine: Up to Date Studies and Future Prospects. *Vaccine* **35**, 5388–5395. doi:[10.1016/j.vaccine.2017.08.034](https://doi.org/10.1016/j.vaccine.2017.08.034) (2017).
131. Klimov, A. *et al. Influenza Virus Titration, Antigenic Characterization, and Serological Methods for Antibody Detection in Influenza Virus* (eds Kawaoka, Y. & Neumann, G.) 25–51 (Humana Press, Totowa, NJ, 2012). doi:[10.1007/978-1-61779-621-0_3](https://doi.org/10.1007/978-1-61779-621-0_3).
132. Cox, R. Correlates of Protection to Influenza Virus, Where Do We Go from Here? *Human Vaccines & Immunotherapeutics* **9**, 405–408. doi:[10.4161/hv.22908](https://doi.org/10.4161/hv.22908) (2013).

133. Pulendran, B., Li, S. & Nakaya, H. I. Systems Vaccinology. *Immunity* **33**, 516–529. doi:[10.1016/j.immuni.2010.10.006](https://doi.org/10.1016/j.immuni.2010.10.006) (2010).
134. Koutsakos, M. *et al.* Circulating T FH Cells, Serological Memory, and Tissue Compartmentalization Shape Human Influenza-Specific B Cell Immunity. *Science Translational Medicine* **10**, eaan8405. doi:[10.1126/scitranslmed.aan8405](https://doi.org/10.1126/scitranslmed.aan8405) (2018).
135. Renegar, K. B., Small, P. A., Boykins, L. G. & Wright, P. F. Role of IgA versus IgG in the Control of Influenza Viral Infection in the Murine Respiratory Tract. *The Journal of Immunology* **173**, 1978–1986. doi:[10.4049/jimmunol.173.3.1978](https://doi.org/10.4049/jimmunol.173.3.1978) (2004).
136. De Gregorio, E. & Rappuoli, R. From Empiricism to Rational Design: A Personal Perspective of the Evolution of Vaccine Development. *Nature Reviews Immunology* **14**, 505–514. doi:[10.1038/nri3694](https://doi.org/10.1038/nri3694) (2014).
137. Nakaya, H. I., Li, S. & Pulendran, B. Systems Vaccinology: Learning to Compute the Behavior of Vaccine Induced Immunity. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **4**, 193–205. doi:[10.1002/wsbm.163](https://doi.org/10.1002/wsbm.163) (2012).
138. Li, S., Nakaya, H. I., Kazmin, D. A., Oh, J. Z. & Pulendran, B. Systems Biological Approaches to Measure and Understand Vaccine Immunity in Humans. *Seminars in Immunology* **25**, 209–218. doi:[10.1016/j.smim.2013.05.003](https://doi.org/10.1016/j.smim.2013.05.003) (2013).
139. Pulendran, B. Systems Vaccinology: Probing Humanity’s Diverse Immune Systems with Vaccines. *Proceedings of the National Academy of Sciences* **111**, 12300–12306. doi:[10.1073/pnas.1400476111](https://doi.org/10.1073/pnas.1400476111) (2014).
140. Hagan, T., Nakaya, H. I., Subramaniam, S. & Pulendran, B. Systems Vaccinology: Enabling Rational Vaccine Design with Systems Biological Approaches. *Vaccine* **33**, 5294–5301. doi:[10.1016/j.vaccine.2015.03.072](https://doi.org/10.1016/j.vaccine.2015.03.072) (2015).
141. Nakaya, H. I. & Pulendran, B. Vaccinology in the Era of High-Throughput Biology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140146–20140146. doi:[10.1098/rstb.2014.0146](https://doi.org/10.1098/rstb.2014.0146) (2015).
142. Davis, M. M. & Tato, C. M. Will Systems Biology Deliver Its Promise and Contribute to the Development of New or Improved Vaccines?: Seeing the Forest Rather than a Few Trees. *Cold Spring Harbor Perspectives in Biology* **10**, a028886. doi:[10.1101/cshperspect.a028886](https://doi.org/10.1101/cshperspect.a028886) (2018).
143. Raeven, R. H. M., van Riet, E., Meiring, H. D., Metz, B. & Kersten, G. F. A. Systems Vaccinology and Big Data in the Vaccine Development Chain. *Immunology* **156**, 33–46. doi:[10.1111/imm.13012](https://doi.org/10.1111/imm.13012) (2019).
144. Siegrist, C.-A. *Vaccine Immunology in Plotkin’s Vaccines* 16–34.e7 (Elsevier, 2018). doi:[10.1016/B978-0-323-35761-6.00002-X](https://doi.org/10.1016/B978-0-323-35761-6.00002-X).
145. Plotkin, S. A. & Gilbert, P. *Correlates of Protection in Plotkin’s Vaccines* 35–40.e4 (Elsevier, 2018). doi:[10.1016/B978-0-323-35761-6.00003-1](https://doi.org/10.1016/B978-0-323-35761-6.00003-1).
146. Tsang, J. S. *et al.* Improving Vaccine-Induced Immunity: Can Baseline Predict Outcome? *Trends in Immunology*. doi:[10.1016/j.it.2020.04.001](https://doi.org/10.1016/j.it.2020.04.001) (2020).

147. Pulendran, B. Learning Immunology from the Yellow Fever Vaccine: Innate Immunity to Systems Vaccinology. *Nature Reviews Immunology* **9**, 741–747. doi:[10.1038/nri2629](https://doi.org/10.1038/nri2629) (2009).
148. Cotugno, N. *et al.* OMIC Technologies and Vaccine Development: From the Identification of Vulnerable Individuals to the Formulation of Invulnerable Vaccines. *Journal of Immunology Research* **2019**, 1–10. doi:[10.1155/2019/8732191](https://doi.org/10.1155/2019/8732191) (2019).
149. Vahey, M. T. *et al.* Expression of Genes Associated with Immunoproteasome Processing of Major Histocompatibility Complex Peptides Is Indicative of Protection with Adjuvanted RTS,S Malaria Vaccine. *The Journal of Infectious Diseases* **201**, 580–589. doi:[10.1086/650310](https://doi.org/10.1086/650310) (2010).
150. Blohmke, C. J., O'Connor, D. & Pollard, A. J. The Use of Systems Biology and Immunological Big Data to Guide Vaccine Development. *Genome Medicine* **7**. doi:[10.1186/s13073-015-0236-1](https://doi.org/10.1186/s13073-015-0236-1) (2015).
151. Zhu, W. *et al.* A Whole Genome Transcriptional Analysis of the Early Immune Response Induced by Live Attenuated and Inactivated Influenza Vaccines in Young Children. *Vaccine* **28**, 2865–2876. doi:[10.1016/j.vaccine.2010.01.060](https://doi.org/10.1016/j.vaccine.2010.01.060) (2010).
152. Cao, R. G. *et al.* Differences in Antibody Responses Between Trivalent Inactivated Influenza Vaccine and Live Attenuated Influenza Vaccine Correlate With the Kinetics and Magnitude of Interferon Signaling in Children. *The Journal of Infectious Diseases* **210**, 224–233. doi:[10.1093/infdis/jiu079](https://doi.org/10.1093/infdis/jiu079) (2014).
153. Bucasas, K. L. *et al.* Early Patterns of Gene Expression Correlate With the Humoral Immune Response to Influenza Vaccination in Humans. *The Journal of Infectious Diseases* **203**, 921–929. doi:[10.1093/infdis/jiq156](https://doi.org/10.1093/infdis/jiq156) (2011).
154. Nakaya, H. I. *et al.* Systems Biology of Vaccination for Seasonal Influenza in Humans. *Nature Immunology* **12**, 786–795. doi:[10.1038/ni.2067](https://doi.org/10.1038/ni.2067) (2011).
155. Tan, Y., Tamayo, P., Nakaya, H., Pulendran, B., Mesirov, J. P. & Haining, W. N. Gene Signatures Related to B-Cell Proliferation Predict Influenza Vaccine-Induced Antibody Response. *European Journal of Immunology* **44**, 285–295. doi:[10.1002/eji.201343657](https://doi.org/10.1002/eji.201343657) (2014).
156. Frasca, D. *et al.* Intrinsic Defects in B Cell Response to Seasonal Influenza Vaccination in Elderly Humans. *Vaccine* **28**, 8077–8084. doi:[10.1016/j.vaccine.2010.10.023](https://doi.org/10.1016/j.vaccine.2010.10.023) (2010).
157. Nakaya, H. I. *et al.* Systems Analysis of Immunity to Influenza Vaccination across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures. *Immunity* **43**, 1186–1198. doi:[10.1016/j.immuni.2015.11.012](https://doi.org/10.1016/j.immuni.2015.11.012) (2015).
158. Nakaya, H. I. *et al.* Systems Biology of Immunity to MF59-Adjuvanted versus Nonadjuvanted Trivalent Seasonal Influenza Vaccines in Early Childhood. *Proceedings of the National Academy of Sciences* **113**, 1853–1858. doi:[10.1073/pnas.1519690113](https://doi.org/10.1073/pnas.1519690113) (2016).
159. HIPC-CHI Signatures Project Team & HIPC-I Consortium. Multicohort Analysis Reveals Baseline Transcriptional Predictors of Influenza Vaccination Responses. *Science Immunology* **2**, eaal4656. doi:[10.1126/sciimmunol.aal4656](https://doi.org/10.1126/sciimmunol.aal4656) (2017).

160. Furman, D. *et al.* Apoptosis and Other Immune Biomarkers Predict Influenza Vaccine Responsiveness. *Molecular Systems Biology* **9**, 659. doi:[10.1038/msb.2013.15](https://doi.org/10.1038/msb.2013.15) (2013).
161. Tsang, J. S. *et al.* Global Analyses of Human Immune Variation Reveal Baseline Predictors of Postvaccination Responses. *Cell* **157**, 499–513. doi:[10.1016/j.cell.2014.03.031](https://doi.org/10.1016/j.cell.2014.03.031) (2014).
162. Sobolev, O. *et al.* Adjuvanted Influenza-H1N1 Vaccination Reveals Lymphoid Signatures of Age-Dependent Early Responses and of Clinical Adverse Events. *Nature Immunology* **17**, 204–213. doi:[10.1038/ni.3328](https://doi.org/10.1038/ni.3328) (2016).
163. Wilkins, A. L. *et al.* AS03- and MF59-Adjuvanted Influenza Vaccines in Children. *Frontiers in Immunology* **8**. doi:[10.3389/fimmu.2017.01760](https://doi.org/10.3389/fimmu.2017.01760) (2017).
164. Tregoning, J. S., Russell, R. F. & Kinnear, E. Adjuvanted Influenza Vaccines. *Human Vaccines & Immunotherapeutics* **14**, 550–564. doi:[10.1080/21645515.2017.1415684](https://doi.org/10.1080/21645515.2017.1415684) (2018).
165. Broadbent, A. J. & Subbarao, K. Influenza Virus Vaccines: Lessons from the 2009 H1N1 Pandemic. *Current Opinion in Virology* **1**, 254–262. doi:[10.1016/j.coviro.2011.08.002](https://doi.org/10.1016/j.coviro.2011.08.002) (2011).
166. Syrjänen, R. K. *et al.* Effectiveness of Pandemic and Seasonal Influenza Vaccines in Preventing Laboratory-Confirmed Influenza in Adults: A Clinical Cohort Study during Epidemic Seasons 2009–2010 and 2010–2011 in Finland. *PLoS ONE* **9** (ed Marques, E. T. A.) e108538. doi:[10.1371/journal.pone.0108538](https://doi.org/10.1371/journal.pone.0108538) (2014).
167. Food and Drug Administration. *Guidance for Industry: Clinical Data Needed to Support the Licensure of Pandemic Influenza Vaccines* (2007), 20.
168. Barnett, A. G. Regression to the Mean: What It Is and How to Deal with It. *International Journal of Epidemiology* **34**, 215–220. doi:[10.1093/ije/dyh299](https://doi.org/10.1093/ije/dyh299) (2004).
169. Senn, S. Francis Galton and Regression to the Mean. *Significance* **8**, 124–126. doi:[10.1111/j.1740-9713.2011.00509.x](https://doi.org/10.1111/j.1740-9713.2011.00509.x) (2011).
170. Tu, Y.-K. & Gilthorpe, M. S. Revisiting the Relation between Change and Initial Value: A Review and Evaluation. *Statistics in Medicine* **26**, 443–457. doi:[10.1002/sim.2538](https://doi.org/10.1002/sim.2538) (2007).
171. Clifton, L. & Clifton, D. A. The Correlation between Baseline Score and Post-Intervention Score, and Its Implications for Statistical Analysis. *Trials* **20**. doi:[10.1186/s13063-018-3108-3](https://doi.org/10.1186/s13063-018-3108-3) (2019).
172. Cohen, J. The Cost of Dichotomization. *Applied Psychological Measurement* **7**, 249–253. doi:[10.1177/014662168300700301](https://doi.org/10.1177/014662168300700301) (1983).
173. Senn, S. *Dichotomania: An Obsessive Compulsive Disorder That Is Badly Affecting the Quality of Analysis of Pharmaceutical Trials in Proceedings of the International Statistical Institute 55th World Statistics Congress* (International Statistical Institute, Sydney, 2005), 14.

174. Altman, D. G. & Royston, P. The Cost of Dichotomising Continuous Variables. *BMJ* **332**, 1080.1. doi:[10.1136/bmj.332.7549.1080](https://doi.org/10.1136/bmj.332.7549.1080) (2006).
175. Fedorov, V., Mannino, F. & Zhang, R. Consequences of Dichotomization. *Pharmaceutical Statistics* **8**, 50–61. doi:[10.1002/pst.331](https://doi.org/10.1002/pst.331) (2009).
176. Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. & Lee, J. J. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaScience* **4**, 7. doi:[10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8) (2015).
177. McCarthy, M. I. *et al.* Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges. *Nature Reviews Genetics* **9**, 356–369. doi:[10.1038/nrg2344](https://doi.org/10.1038/nrg2344) (2008).
178. Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P. & Zondervan, K. T. Data Quality Control in Genetic Case-Control Association Studies. *Nature Protocols* **5**, 1564–73. doi:[10.1038/nprot.2010.116](https://doi.org/10.1038/nprot.2010.116) (2010).
179. Marees, A. T. *et al.* A Tutorial on Conducting Genome-Wide Association Studies: Quality Control and Statistical Analysis. *International Journal of Methods in Psychiatric Research* **27**, e1608. doi:[10.1002/mpr.1608](https://doi.org/10.1002/mpr.1608) (2018).
180. De Lange, K. M. *et al.* Genome-Wide Association Study Implicates Immune Activation of Multiple Integrin Genes in Inflammatory Bowel Disease. *Nature Genetics* **49**, 256–261. doi:[10.1038/ng.3760](https://doi.org/10.1038/ng.3760) (2017).
181. Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M. & Chen, W.-M. Robust Relationship Inference in Genome-Wide Association Studies. *Bioinformatics* **26**, 2867–2873. doi:[10.1093/bioinformatics/btq559](https://doi.org/10.1093/bioinformatics/btq559) (2010).
182. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using Probabilistic Estimation of Expression Residuals (PEER) to Obtain Increased Power and Interpretability of Gene Expression Analyses. *Nature protocols* **7**, 500–507. doi:[10.1038/nprot.2011.457](https://doi.org/10.1038/nprot.2011.457) (2012).
183. Brown, B. C., Bray, N. L. & Pachter, L. Expression Reflects Population Structure. *PLOS Genetics* **14** (ed Di Rienzo, A.) e1007841. doi:[10.1371/journal.pgen.1007841](https://doi.org/10.1371/journal.pgen.1007841) (2018).
184. The International HapMap 3 Consortium. Integrating Common and Rare Genetic Variation in Diverse Human Populations. *Nature* **467**, 52–58. doi:[10.1038/nature09298](https://doi.org/10.1038/nature09298) (2010).
185. Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. *Nature Genetics* **38**, 904–909. doi:[10.1038/ng1847](https://doi.org/10.1038/ng1847) (2006).
186. Abdellaoui, A. *et al.* Population Structure, Migration, and Diversifying Selection in the Netherlands. *European Journal of Human Genetics* **21**, 1277–1285. doi:[10.1038/ejhg.2013.48](https://doi.org/10.1038/ejhg.2013.48) (2013).
187. Privé, F., Luu, K., Blum, M. G. B., McGrath, J. J. & Vilhjálmsson, B. J. Efficient Toolkit Implementing Best Practices for Principal Component Analysis of Population Genetic Data. *Bioinformatics* **36** (ed Schwartz, R.) 4449–4457. doi:[10.1093/bioinformatics/btaa520](https://doi.org/10.1093/bioinformatics/btaa520) (2020).

188. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New Approaches to Population Stratification in Genome-Wide Association Studies. *Nature Reviews Genetics* **11**, 459–463. doi:[10.1038/nrg2813](https://doi.org/10.1038/nrg2813) (2010).
189. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genetics* **2**, e190. doi:[10.1371/journal.pgen.0020190](https://doi.org/10.1371/journal.pgen.0020190) (2006).
190. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* **32**, btv566. doi:[10.1093/bioinformatics/btv566](https://doi.org/10.1093/bioinformatics/btv566) (2015).
191. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* **32**, 3047–3048. doi:[10.1093/bioinformatics/btw354](https://doi.org/10.1093/bioinformatics/btw354) (2016).
192. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nature Methods* **14**, 417–419. doi:[10.1038/nmeth.4197](https://doi.org/10.1038/nmeth.4197) (2017).
193. Soneson, C., Love, M. I. & Robinson, M. D. Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences. *F1000Research* **4**, 1521. doi:[10.12688/f1000research.7563.2](https://doi.org/10.12688/f1000research.7563.2) (2016).
194. Love, M. I., Soneson, C. & Patro, R. Swimming Downstream: Statistical Analysis of Differential Transcript Usage Following Salmon Quantification. *F1000Research* **7**, 952. doi:[10.12688/f1000research.15398.3](https://doi.org/10.12688/f1000research.15398.3) (2018).
195. Liu, Y., Zhou, J. & White, K. P. RNA-Seq Differential Expression Studies: More Sequence or More Replication? *Bioinformatics* **30**, 301–304. doi:[10.1093/bioinformatics/btt688](https://doi.org/10.1093/bioinformatics/btt688) (2014).
196. Conesa, A. *et al.* A Survey of Best Practices for RNA-Seq Data Analysis. *Genome Biology* **17**, 1–19. doi:[10.1186/s13059-016-0881-8](https://doi.org/10.1186/s13059-016-0881-8) (2016).
197. Zhao, S., Zhang, Y., Gamini, R., Zhang, B. & von Schack, D. Evaluation of Two Main RNA-Seq Approaches for Gene Quantification in Clinical RNA Sequencing: polyA+ Selection versus rRNA Depletion. *Scientific Reports* **8**. doi:[10.1038/s41598-018-23226-4](https://doi.org/10.1038/s41598-018-23226-4) (2018).
198. Min, J. L. *et al.* Variability of Gene Expression Profiles in Human Blood and Lymphoblastoid Cell Lines. *BMC Genomics* **11**, 96. doi:[10.1186/1471-2164-11-96](https://doi.org/10.1186/1471-2164-11-96) (2010).
199. Chen, Y., Lun, A. T. L. & Smyth, G. K. From Reads to Genes to Pathways: Differential Expression Analysis of RNA-Seq Experiments Using Rsubread and the edgeR Quasi-Likelihood Pipeline. *F1000Research* **5**, 1438. doi:[10.12688/f1000research.8987.2](https://doi.org/10.12688/f1000research.8987.2) (2016).
200. Law, C. W. *et al.* RNA-Seq Analysis Is Easy as 1-2-3 with Limma, Glimma and edgeR. *F1000Research* **5**, 1408. doi:[10.12688/f1000research.9005.3](https://doi.org/10.12688/f1000research.9005.3) (2018).
201. Robinson, M. D. & Oshlack, A. A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data. *Genome Biology* **11**, R25. doi:[10.1186/gb-2010-11-3-r25](https://doi.org/10.1186/gb-2010-11-3-r25) (2010).

202. Evans, C., Hardin, J. & Stoebel, D. M. Selecting Between-Sample RNA-Seq Normalization Methods from the Perspective of Their Assumptions. *Briefings in Bioinformatics* **19**, 776–792. doi:[10.1093/bib/bbx008](https://doi.org/10.1093/bib/bbx008) (2018).
203. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **26**, 139–140. doi:[10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616) (2010).
204. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biology* **15**, 1–17 (2014).
205. Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression. *Bioinformatics* **18**, S96–S104. doi:[10.1093/bioinformatics/18.suppl_1.S96](https://doi.org/10.1093/bioinformatics/18.suppl_1.S96) (Suppl 1 2002).
206. Miller, J. A. *et al.* Strategies for Aggregating Gene Expression Data: The collapseRows R Function. *BMC Bioinformatics* **12**, 322. doi:[10.1186/1471-2105-12-322](https://doi.org/10.1186/1471-2105-12-322) (2011).
207. Ritchie, M. E. *et al.* Empirical Array Quality Weights in the Analysis of Microarray Data. *BMC Bioinformatics* **7**, 261. doi:[10.1186/1471-2105-7-261](https://doi.org/10.1186/1471-2105-7-261) (2006).
208. Robinson, D. G., Wang, J. Y. & Storey, J. D. A Nested Parallel Experiment Demonstrates Differences in Intensity-Dependence between RNA-Seq and Microarrays. *Nucleic Acids Research*, gkv636. doi:[10.1093/nar/gkv636](https://doi.org/10.1093/nar/gkv636) (2015).
209. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nature Reviews Genetics* **10**, 57–63. doi:[10.1038/nrg2484](https://doi.org/10.1038/nrg2484) (2009).
210. Ma, T., Liang, F., Oesterreich, S. & Tseng, G. C. A Joint Bayesian Model for Integrating Microarray and RNA Sequencing Transcriptomic Data. *Journal of Computational Biology* **24**, 647–662. doi:[10.1089/cmb.2017.0056](https://doi.org/10.1089/cmb.2017.0056) (2017).
211. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* **8**, 118–127. doi:[10.1093/biostatistics/kxj037](https://doi.org/10.1093/biostatistics/kxj037) (2007).
212. Chen, C. *et al.* Removing Batch Effects in Analysis of Expression Microarray Data: An Evaluation of Six Batch Adjustment Methods. *PLoS ONE* **6** (ed Kliebenstein, D.) e17238. doi:[10.1371/journal.pone.0017238](https://doi.org/10.1371/journal.pone.0017238) (2011).
213. Espín-Pérez, A., Portier, C., Chadeau-Hyam, M., van Veldhoven, K., Kleinjans, J. C. S. & de Kok, T. M. C. M. Comparison of Statistical Methods and the Use of Quality Control Samples for Batch Effect Correction in Human Transcriptome Data. *PLOS ONE* **13** (ed Krishnan, V. V.) e0202947. doi:[10.1371/journal.pone.0202947](https://doi.org/10.1371/journal.pone.0202947) (2018).
214. Zhang, Y., Jenkins, D. F., Manimaran, S. & Johnson, W. E. Alternative Empirical Bayes Models for Adjusting for Batch Effects in Genomic Studies. *BMC Bioinformatics* **19**. doi:[10.1186/s12859-018-2263-6](https://doi.org/10.1186/s12859-018-2263-6) (2018).

215. Nygaard, V., Rødland, E. A. & Hovig, E. Methods That Remove Batch Effects While Retaining Group Differences May Lead to Exaggerated Confidence in Downstream Analyses. *Biostatistics*, kxv027. doi:[10.1093/biostatistics/kxv027](https://doi.org/10.1093/biostatistics/kxv027) (January 2015).
216. Ritchie, M. E. *et al.* Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Research* **43**, e47–e47. doi:[10.1093/nar/gkv007](https://doi.org/10.1093/nar/gkv007) (2015).
217. Soneson, C. & Delorenzi, M. A Comparison of Methods for Differential Expression Analysis of RNA-Seq Data. *BMC Bioinformatics* **14**. doi:[10.1186/1471-2105-14-91](https://doi.org/10.1186/1471-2105-14-91) (2013).
218. Cohn, L. D. & Becker, B. J. How Meta-Analysis Increases Statistical Power. *Psychological Methods* **8**, 243–253. doi:[10.1037/1082-989X.8.3.243](https://doi.org/10.1037/1082-989X.8.3.243) (2003).
219. Borenstein, M., Hedges, L. V., Higgins, J. P. & Rothstein, H. R. A Basic Introduction to Fixed-Effect and Random-Effects Models for Meta-Analysis. *Research Synthesis Methods* **1**, 97–111. doi:[10.1002/jrsm.12](https://doi.org/10.1002/jrsm.12) (2010).
220. Röver, C. *Bayesian Random-Effects Meta-Analysis Using the Bayesmeta R Package* <http://arxiv.org/abs/1711.08683> (2018).
221. Higgins, J. P. T., Thompson, S. G. & Spiegelhalter, D. J. A Re-Evaluation of Random-Effects Meta-Analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**, 137–159. doi:[10.1111/j.1467-985X.2008.00552.x](https://doi.org/10.1111/j.1467-985X.2008.00552.x) (2009).
222. Yuen, T. Accuracy and Calibration of Commercial Oligonucleotide and Custom cDNA Microarrays. *Nucleic Acids Research* **30**, 48e–48. doi:[10.1093/nar/30.10.e48](https://doi.org/10.1093/nar/30.10.e48) (2002).
223. Draghici, S., Khatri, P., Eklund, A. & Szallasi, Z. Reliability and Reproducibility Issues in DNA Microarray Measurements. *Trends in Genetics* **22**, 101–109. doi:[10.1016/j.tig.2005.12.005](https://doi.org/10.1016/j.tig.2005.12.005) (2006).
224. Bender, R. *et al.* Methods for Evidence Synthesis in the Case of Very Few Studies. *Research Synthesis Methods*. doi:[10.1002/jrsm.1297](https://doi.org/10.1002/jrsm.1297) (2018).
225. Gonnermann, A., Framke, T., Großhennig, A. & Koch, A. No Solution yet for Combining Two Independent Studies in the Presence of Heterogeneity. *Statistics in Medicine* **34**, 2476–2480. doi:[10.1002/sim.6473](https://doi.org/10.1002/sim.6473) (2015).
226. Veroniki, A. A. *et al.* Methods to Estimate the Between-Study Variance and Its Uncertainty in Meta-Analysis. *Research Synthesis Methods* **7**, 55–79. doi:[10.1002/jrsm.1164](https://doi.org/10.1002/jrsm.1164) (2016).
227. Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A. & Liu, J. A Nondegenerate Penalized Likelihood Estimator for Variance Parameters in Multilevel Models. *Psychometrika* **78**, 685–709. doi:[10.1007/s11336-013-9328-2](https://doi.org/10.1007/s11336-013-9328-2) (2013).
228. Friede, T., Röver, C., Wandel, S. & Neuenschwander, B. Meta-Analysis of Few Small Studies in Orphan Diseases. *Research Synthesis Methods* **8**, 79–91. doi:[10.1002/jrsm.1217](https://doi.org/10.1002/jrsm.1217) (2017).
229. Seide, S. E., Röver, C. & Friede, T. Likelihood-Based Random-Effects Meta-Analysis with Few Studies: Empirical and Simulation Studies. *BMC Medical Research Methodology* **19**. doi:[10.1186/s12874-018-0618-3](https://doi.org/10.1186/s12874-018-0618-3) (2019).

230. Gelman, A. Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper). *Bayesian Analysis* **1**, 515–534. doi:[10.1214/06-BA117A](https://doi.org/10.1214/06-BA117A) (2006).
231. Pullenayegum, E. M. An Informed Reference Prior for Between-Study Heterogeneity in Meta-Analyses of Binary Outcomes: Prior for between-Study Heterogeneity. *Statistics in Medicine* **30**, 3082–3094. doi:[10.1002/sim.4326](https://doi.org/10.1002/sim.4326) (2011).
232. Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G. & Higgins, J. P. T. Predictive Distributions for Between-Study Heterogeneity and Simple Methods for Their Application in Bayesian Meta-Analysis. *Statistics in Medicine* **34**, 984–998. doi:[10.1002/sim.6381](https://doi.org/10.1002/sim.6381) (2015).
233. Higgins, J. P. T. & Whitehead, A. Borrowing Strength from External Trials in a Meta-Analysis. *Statistics in Medicine* **15**, 2733–2749. doi:[10.1002/\(SICI\)1097-0258\(19961230\)15:24<2733::AID-SIM562>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0) (1996).
234. Viechtbauer, W. Conducting Meta-Analyses in *R* with the **Metafor** Package. *Journal of Statistical Software* **36**. doi:[10.18637/jss.v036.i03](https://doi.org/10.18637/jss.v036.i03) (2010).
235. Delignette-Muller, M. L. & Dutang, C. **Fitdistrplus** : An *R* Package for Fitting Distributions. *Journal of Statistical Software* **64**. doi:[10.18637/jss.v064.i04](https://doi.org/10.18637/jss.v064.i04) (2015).
236. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-Tailed Prior Distributions for Sequence Count Data: Removing the Noise and Preserving Large Differences. *Bioinformatics* **35** (ed Stegle, O.) 2084–2092. doi:[10.1093/bioinformatics/bty895](https://doi.org/10.1093/bioinformatics/bty895) (2019).
237. Love, M. I., Huber, W. & Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biology* **15**, 550. doi:[10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8) (2014).
238. Stephens, M. False Discovery Rates: A New Deal. *Biostatistics*, kxw041. doi:[10.1093/biostatistics/kxw041](https://doi.org/10.1093/biostatistics/kxw041) (2016).
239. Chaussabel, D. *et al.* A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus. *Immunity* **29**, 150–164. doi:[10.1016/j.jimmuni.2008.05.012](https://doi.org/10.1016/j.jimmuni.2008.05.012) (2008).
240. Li, S. *et al.* Molecular Signatures of Antibody Responses Derived from a Systems Biology Study of Five Human Vaccines. *Nature Immunology* **15**, 195–204. doi:[10.1038/ni.2789](https://doi.org/10.1038/ni.2789) (2013).
241. Weiner 3rd, J. & Domaszewska, T. Tmod: An R Package for General and Multivariate Enrichment Analysis. doi:[10.7287/peerj.preprints.2420v1](https://doi.org/10.7287/peerj.preprints.2420v1) (2016).
242. Zyla, J., Marczyk, M., Domaszewska, T., Kaufmann, S. H. E., Polanska, J. & Weiner, J. Gene Set Enrichment for Reproducible Science: Comparison of CERNO and Eight Other Algorithms. *Bioinformatics* **35** (ed Wren, J.) 5146–5154. doi:[10.1093/bioinformatics/btz447](https://doi.org/10.1093/bioinformatics/btz447) (2019).

243. Bin, L., Li, X., Feng, J., Richers, B. & Leung, D. Y. M. Ankyrin Repeat Domain 22 Mediates Host Defense Against Viral Infection Through STING Signaling Pathway. *The Journal of Immunology* **196**, 2014 LP –2014 (1 Supplement 2016).
244. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annual Review of Immunology* **32**, 513–545. doi:[10.1146/annurev-immunol-032713-120231](https://doi.org/10.1146/annurev-immunol-032713-120231) (2014).
245. Ogawa, K. *et al.* A Novel Serum Protein That Is Selectively Produced by Cytotoxic Lymphocytes. *The Journal of Immunology* **166**, 6404–6412. doi:[10.4049/jimmunol.166.10.6404](https://doi.org/10.4049/jimmunol.166.10.6404) (2001).
246. McCarron, M. J., Park, P. W. & Fooksman, D. R. CD138 Mediates Selection of Mature Plasma Cells by Regulating Their Survival. *Blood* **129**, 2749–2759. doi:[10.1182/blood-2017-01-761643](https://doi.org/10.1182/blood-2017-01-761643) (2017).
247. Martincic, K., Alkan, S. A., Cheatle, A., Borghesi, L. & Milcarek, C. Transcription Elongation Factor ELL2 Directs Immunoglobulin Secretion in Plasma Cells by Stimulating Altered RNA Processing. *Nature Immunology* **10**, 1102–1109. doi:[10.1038/ni.1786](https://doi.org/10.1038/ni.1786) (2009).
248. Morel, S. *et al.* Adjuvant System AS03 Containing α -Tocopherol Modulates Innate Immune Response and Leads to Improved Adaptive Immunity. *Vaccine* **29**, 2461–2473. doi:[10.1016/j.vaccine.2011.01.011](https://doi.org/10.1016/j.vaccine.2011.01.011) (2011).
249. Querec, T. D. *et al.* Systems Biology Approach Predicts Immunogenicity of the Yellow Fever Vaccine in Humans. *Nature Immunology* **10**, 116–125. doi:[10.1038/ni.1688](https://doi.org/10.1038/ni.1688) (2009).
250. Mitchell, L. A., Henderson, A. J. & Dow, S. W. Suppression of Vaccine Immunity by Inflammatory Monocytes. *The Journal of Immunology* **189**, 5612–5621. doi:[10.4049/jimmunol.1202151](https://doi.org/10.4049/jimmunol.1202151) (2012).
251. Mohanty, S. *et al.* Prolonged Proinflammatory Cytokine Production in Monocytes Modulated by Interleukin 10 After Influenza Vaccination in Older Adults. *Journal of Infectious Diseases* **211**, 1174–1184. doi:[10.1093/infdis/jiu573](https://doi.org/10.1093/infdis/jiu573) (2015).
252. Mooney, M., McWeeney, S. & Sékaly, R.-P. Systems Immunogenetics of Vaccines. *Seminars in Immunology* **25**, 124–129. doi:[10.1016/j.smim.2013.06.003](https://doi.org/10.1016/j.smim.2013.06.003) (2013).
253. O'Connor, D. & Pollard, A. J. Characterizing Vaccine Responses Using Host Genomic and Transcriptomic Analysis. *Clinical Infectious Diseases* **57**, 860–869. doi:[10.1093/cid/cit373](https://doi.org/10.1093/cid/cit373) (2013).
254. Newport, M. J. The Genetic Regulation of Infant Immune Responses to Vaccination. *Frontiers in Immunology* **6**. doi:[10.3389/fimmu.2015.00018](https://doi.org/10.3389/fimmu.2015.00018) (2015).
255. Brodin, P. *et al.* Variation in the Human Immune System Is Largely Driven by Non-Heritable Influences. *Cell* **160**, 37–47. doi:[10.1016/j.cell.2014.12.020](https://doi.org/10.1016/j.cell.2014.12.020) (2015).

256. Mentzer, A. J., O'Connor, D., Pollard, A. J. & Hill, A. V. S. Searching for the Human Genetic Factors Standing in the Way of Universally Effective Vaccines. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140341–20140341. doi:[10.1098/rstb.2014.0341](https://doi.org/10.1098/rstb.2014.0341) (2015).
257. Linnik, J. E. & Egli, A. Impact of Host Genetic Polymorphisms on Vaccine Induced Antibody Response. *Human Vaccines & Immunotherapeutics* **12**, 907–915. doi:[10.1080/21645515.2015.1119345](https://doi.org/10.1080/21645515.2015.1119345) (2016).
258. Gelder, C. M. *et al.* Associations between Human Leukocyte Antigens and Nonresponsiveness to Influenza Vaccine. *The Journal of Infectious Diseases* **185**, 114–117. doi:[10.1086/338014](https://doi.org/10.1086/338014) (2002).
259. Moss, A. J. *et al.* Correlation between Human Leukocyte Antigen Class II Alleles and HAI Titers Detected Post-Influenza Vaccination. *PLoS ONE* **8** (ed Sambhara, S.) e71376. doi:[10.1371/journal.pone.0071376](https://doi.org/10.1371/journal.pone.0071376) (2013).
260. Poland, G. A., Ovsyannikova, I. G. & Jacobson, R. M. Immunogenetics of Seasonal Influenza Vaccine Response. *Vaccine* **26**, D35–D40. doi:[10.1016/j.vaccine.2008.07.065](https://doi.org/10.1016/j.vaccine.2008.07.065) (2008).
261. Egli, A. *et al.* IL-28B Is a Key Regulator of B- and T-Cell Vaccine Responses against Influenza. *PLoS Pathogens* **10** (ed Gale, M.) e1004556. doi:[10.1371/journal.ppat.1004556](https://doi.org/10.1371/journal.ppat.1004556) (2014).
262. Avnir, Y. *et al.* IGHV1-69 Polymorphism Modulates Anti-Influenza Antibody Repertoires, Correlates with IGHV Utilization Shifts and Varies by Ethnicity. *Scientific Reports* **6**, 20842. doi:[10.1038/srep20842](https://doi.org/10.1038/srep20842) (2016).
263. Millstein, J., Zhang, B., Zhu, J. & Schadt, E. E. Disentangling Molecular Relationships with a Causal Inference Test. *BMC Genetics* **10**. doi:[10.1186/1471-2156-10-23](https://doi.org/10.1186/1471-2156-10-23) (2009).
264. Astle, W. & Balding, D. J. Population Structure and Cryptic Relatedness in Genetic Association Studies. *Statistical Science* **24**, 451–471. doi:[10.1214/09-STS307](https://doi.org/10.1214/09-STS307) (2009).
265. Sillanpää, M. J. Overview of Techniques to Account for Confounding Due to Population Stratification and Cryptic Relatedness in Genomic Data Association Analyses. *Heredity* **106**, 511–519. doi:[10.1038/hdy.2010.91](https://doi.org/10.1038/hdy.2010.91) (2011).
266. Sul, J. H., Martin, L. S. & Eskin, E. Population Structure in Genetic Studies: Confounding Factors and Mixed Models. *PLOS Genetics* **14** (ed Barsh, G. S.) e1007309. doi:[10.1371/journal.pgen.1007309](https://doi.org/10.1371/journal.pgen.1007309) (2018).
267. Golan, D., Rosset, S. & Lin, D.-Y. *Mixed Models for Case-Control Genome-Wide Association Studies: Major Challenges and Partial Solutions* in Borga, Ø., Breslow, N. E., Chatterjee, N., Gail, M. H., Scott, A. & Wild, C. J. *Handbook of Statistical Methods for Case-Control Studies* (eds Borga, Ø., Breslow, N., Chatterjee, N., Gail, M. H., Scott, A. & Wild, C. J.) 1st ed., 495–514 (Chapman and Hall/CRC, 2018). doi:[10.1201/9781315154084-27](https://doi.org/10.1201/9781315154084-27).
268. Vilhjálmsson, B. J. & Nordborg, M. The Nature of Confounding in Genome-Wide Association Studies. *Nature Reviews Genetics* **14**, 1–2. doi:[10.1038/nrg3382](https://doi.org/10.1038/nrg3382) (2013).

269. Eu-ahsunthornwattana, J. *et al.* Comparison of Methods to Account for Relatedness in Genome-Wide Association Studies with Family-Based Data. *PLoS Genetics* **10** (ed Abecasis, G. R.) e1004445. doi:[10.1371/journal.pgen.1004445](https://doi.org/10.1371/journal.pgen.1004445) (2014).
270. Speed, D., Cai, N., Johnson, M. R., Nejentsev, S. & Balding, D. J. Reevaluation of SNP Heritability in Complex Human Traits. *Nature Genetics* **49**, 986–992. doi:[10.1038/ng.3865](https://doi.org/10.1038/ng.3865) (7 2017).
271. Wang, J. Marker-Based Estimates of Relatedness and Inbreeding Coefficients: An Assessment of Current Methods. *Journal of Evolutionary Biology* **27**, 518–530. doi:[10.1111/jeb.12315](https://doi.org/10.1111/jeb.12315) (2014).
272. Widmer, C. *et al.* Further Improvements to Linear Mixed Models for Genome-Wide Association Studies. *Scientific Reports* **4**. doi:[10.1038/srep06874](https://doi.org/10.1038/srep06874) (2015).
273. Lippert, C., Casale, F. P., Rakitsch, B. & Stegle, O. LIMIX: Genetic Analysis of Multiple Traits. doi:[10.1101/003905](https://doi.org/10.1101/003905) (2014).
274. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved Heritability Estimation from Genome-Wide SNPs. *The American Journal of Human Genetics* **91**, 1011–1021. doi:[10.1016/j.ajhg.2012.10.010](https://doi.org/10.1016/j.ajhg.2012.10.010) (2012).
275. Maranville, J. C. *et al.* Interactions between Glucocorticoid Treatment and Cis-Regulatory Polymorphisms Contribute to Cellular Response Phenotypes. *PLoS Genetics* **7** (ed Gibson, G.) e1002162. doi:[10.1371/journal.pgen.1002162](https://doi.org/10.1371/journal.pgen.1002162) (2011).
276. Ackermann, M., Sikora-Wohlfeld, W. & Beyer, A. Impact of Natural Genetic Variation on Gene Expression Dynamics. *PLoS Genetics* **9** (ed Wells, C. A.) e1003514. doi:[10.1371/journal.pgen.1003514](https://doi.org/10.1371/journal.pgen.1003514) (2013).
277. Shpak, M. *et al.* An eQTL Analysis of the Human Glioblastoma Multiforme Genome. *Genomics* **103**, 252–263. doi:[10.1016/j.ygeno.2014.02.005](https://doi.org/10.1016/j.ygeno.2014.02.005) (2014).
278. Allison, P. D. Change Scores as Dependent Variables in Regression Analysis. *Sociological Methodology* **20**, 93. doi:[10.2307/271083](https://doi.org/10.2307/271083) (1990).
279. Clogg, C. C., Petkova, E. & Haritou, A. Statistical Methods for Comparing Regression Coefficients Between Models. *The American Journal of Sociology* **100**, 1261–1293 (1995).
280. Flutre, T., Wen, X., Pritchard, J. & Stephens, M. A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. *PLOS Genet* **9**, e1003486. doi:[10.1371/journal.pgen.1003486](https://doi.org/10.1371/journal.pgen.1003486) (2013).
281. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible Statistical Methods for Estimating and Testing Effects in Genomic Studies with Multiple Conditions. *Nature Genetics*. doi:[10.1038/s41588-018-0268-8](https://doi.org/10.1038/s41588-018-0268-8) (2018).
282. Li, G., Jima, D., Wright, F. A. & Nobel, A. B. HT-eQTL: Integrative Expression Quantitative Trait Loci Analysis in a Large Number of Human Tissues. *BMC Bioinformatics* **19**. doi:[10.1186/s12859-018-2088-3](https://doi.org/10.1186/s12859-018-2088-3) (2018).

283. Stephens, M. A Unified Framework for Association Analysis with Multiple Related Phenotypes. *PLoS ONE* **8** (ed Emmert-Streib, F.) e65245. doi:[10.1371/journal.pone.0065245](https://doi.org/10.1371/journal.pone.0065245) (2013).
284. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively Identifying eQTLs from Multiple Tissues by Combining Mixed Model and Meta-Analytic Approaches. *PLoS Genetics* **9** (ed Schork, N. J.) e1003491. doi:[10.1371/journal.pgen.1003491](https://doi.org/10.1371/journal.pgen.1003491) (2013).
285. Han, B. & Eskin, E. Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-Wide Association Studies. *The American Journal of Human Genetics* **88**, 586–598. doi:[10.1016/j.ajhg.2011.04.014](https://doi.org/10.1016/j.ajhg.2011.04.014) (2011).
286. Lewin, A. *et al.* MT-HESS: An Efficient Bayesian Approach for Simultaneous Association Detection in OMICS Datasets, with Application to eQTL Mapping in Multiple Tissues. *Bioinformatics* **32**, 523–532. doi:[10.1093/bioinformatics/btv568](https://doi.org/10.1093/bioinformatics/btv568) (2016).
287. Li, G., Shabalin, A. A., Rusyn, I., Wright, F. A. & Nobel, A. B. An Empirical Bayes Approach for Multiple Tissue eQTL Analysis. *Biostatistics (Oxford, England)* **19**, 391–406. doi:[10.1093/biostatistics/kxx048](https://doi.org/10.1093/biostatistics/kxx048) (2018).
288. Duong, D. *et al.* Applying Meta-Analysis to Genotype-Tissue Expression Data from Multiple Tissues to Identify eQTLs and Increase the Number of eGenes. *Bioinformatics* **33**, i67–i74. doi:[10.1093/bioinformatics/btx227](https://doi.org/10.1093/bioinformatics/btx227) (2017).
289. McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. & Lin, X. Operating Characteristics of the Rank-based Inverse Normal Transformation for Quantitative Trait Analysis in Genome-wide Association Studies. *Biometrics*. doi:[10.1111/biom.13214](https://doi.org/10.1111/biom.13214) (2020).
290. Beasley, T. M., Erickson, S. & Allison, D. B. Rank-Based Inverse Normal Transformations Are Increasingly Used, But Are They Merited? *Behavior Genetics* **39**, 580–595. doi:[10.1007/s10519-009-9281-0](https://doi.org/10.1007/s10519-009-9281-0) (2009).
291. Qi, T. *et al.* Identifying Gene Targets for Brain-Related Traits Using Transcriptomic and Methylation Data from Blood. *Nature Communications* **9**. doi:[10.1038/s41467-018-04558-1](https://doi.org/10.1038/s41467-018-04558-1) (2018).
292. Loh, P.-R. *et al.* Reference-Based Phasing Using the Haplotype Reference Consortium Panel. *Nature Genetics* **48**, 1443–1448. doi:[10.1038/ng.3679](https://doi.org/10.1038/ng.3679) (2016).
293. Durbin, R. Efficient Haplotype Matching and Storage Using the Positional Burrows-Wheeler Transform (PBWT). *Bioinformatics* **30**, 1266–1272. doi:[10.1093/bioinformatics/btu014](https://doi.org/10.1093/bioinformatics/btu014) (2014).
294. McCarthy, S. *et al.* A Reference Panel of 64,976 Haplotypes for Genotype Imputation. *Nature Genetics* **48**, 1279–1283. doi:[10.1038/ng.3643](https://doi.org/10.1038/ng.3643) (2016).
295. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P. & Wang, L. CrossMap: A Versatile Tool for Coordinate Conversion between Genome Assemblies. *Bioinformatics* **30**, 1006–1007. doi:[10.1093/bioinformatics/btt730](https://doi.org/10.1093/bioinformatics/btt730) (2014).

296. Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E. & Heckerman, D. Improved Linear Mixed Models for Genome-Wide Association Studies. *Nature Methods* **9**, 525–526. doi:[10.1038/nmeth.2037](https://doi.org/10.1038/nmeth.2037) (2012).
297. Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I. & Heckerman, D. FaST Linear Mixed Models for Genome-Wide Association Studies. *Nature Methods* **8**, 833–835. doi:[10.1038/nmeth.1681](https://doi.org/10.1038/nmeth.1681) (2011).
298. Aran, D., Hu, Z. & Butte, A. J. xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biology* **18**. doi:[10.1186/s13059-017-1349-1](https://doi.org/10.1186/s13059-017-1349-1) (2017).
299. Kleiveland, C. R. *Peripheral Blood Mononuclear Cells in The Impact of Food Bioactives on Health* (eds Verhoeckx, K. *et al.*) 161–167 (Springer International Publishing, Cham, 2015). doi:[10.1007/978-3-319-16104-4_15](https://doi.org/10.1007/978-3-319-16104-4_15).
300. Van der Wijst, M. G. P. *et al.* Single-Cell RNA Sequencing Identifies Celltype-Specific Cis-eQTLs and Co-Expression QTLs. *Nature Genetics* **50**, 493–497. doi:[10.1038/s41588-018-0089-9](https://doi.org/10.1038/s41588-018-0089-9) (2018).
301. Maddala, G. S. *Introduction to Econometrics* 2nd ed. 631 pp. (Macmillan Pub. Co. ; Maxwell Macmillan Canada ; Maxwell Macmillan International, New York : Toronto : New York, 1992).
302. Kanyongo, G. Y. The Influence of Reliability on Four Rules for Determining the Number of Components to Retain. *Journal of Modern Applied Statistical Methods* **5**, 332–343. doi:[10.22237/jmasm/1162353960](https://doi.org/10.22237/jmasm/1162353960) (2005).
303. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415–1429.e19. doi:[10.1016/j.cell.2016.10.042](https://doi.org/10.1016/j.cell.2016.10.042) (2016).
304. Stekhoven, D. J. & Bühlmann, P. MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data. *Bioinformatics* **28**, 112–118. doi:[10.1093/bioinformatics/btr597](https://doi.org/10.1093/bioinformatics/btr597) (2012).
305. Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, False Discovery Rate and Winner’s Curse in eQTL Studies. *Nucleic Acids Research* **46**, e133–e133. doi:[10.1093/nar/gky780](https://doi.org/10.1093/nar/gky780) (2018).
306. Schenker, N. & Gentleman, J. F. On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals. *The American Statistician* **55**, 182–186 (2001).
307. Gelman, A. & Stern, H. The Difference Between “Significant” and “Not Significant” Is Not Itself Statistically Significant. *The American Statistician* **60**, 328–331. doi:[10.1198/00313006X152649](https://doi.org/10.1198/00313006X152649) (2006).
308. Shim, H. *et al.* A Multivariate Genome-Wide Association Analysis of 10 LDL Subfractions, and Their Response to Statin Treatment, in 1868 Caucasians. *PLOS ONE* **10** (ed Aspichueta, P.) e0120758. doi:[10.1371/journal.pone.0120758](https://doi.org/10.1371/journal.pone.0120758) (2015).

309. Storey, J. D. & Tibshirani, R. Statistical Significance for Genomewide Studies. *Proceedings of the National Academy of Sciences* **100**, 9440–9445. doi:[10.1073/pnas.1530509100](https://doi.org/10.1073/pnas.1530509100) (2003).
310. Mikucka, M., Sarracino, F. & Dubrow, J. *Costs and Benefits of Including or Omitting Interaction Terms: A Monte Carlo Simulation* 9 (The Ohio State University and the Polish Academy of Sciences, 2015).
311. Kooperberg, C. & LeBlanc, M. Increasing the Power of Identifying Gene × Gene Interactions in Genome-Wide Association Studies. *Genetic Epidemiology* **32**, 255–263. doi:[10.1002/gepi.20300](https://doi.org/10.1002/gepi.20300) (2008).
312. Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H. & Soria, J. M. Lme4qtl: Linear Mixed Models with Flexible Covariance Structure for Genetic Studies of Related Individuals. *BMC Bioinformatics* **19**, 68. doi:[10.1186/s12859-018-2057-x](https://doi.org/10.1186/s12859-018-2057-x) (2018).
313. Raudvere, U. *et al.* G:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update). *Nucleic Acids Research* **47**, W191–W198. doi:[10.1093/nar/gkz369](https://doi.org/10.1093/nar/gkz369) (2019).
314. Kerimov, N. *et al.* eQTL Catalogue: A Compendium of Uniformly Processed Human Gene Expression and Splicing QTLs. *bioRxiv*. doi:[10.1101/2020.01.29.924266](https://doi.org/10.1101/2020.01.29.924266) (2020).
315. Lawrence, M., Gentleman, R. & Carey, V. Rtracklayer: An R Package for Interfacing with Genome Browsers. *Bioinformatics* **25**, 1841–1842. doi:[10.1093/bioinformatics/btp328](https://doi.org/10.1093/bioinformatics/btp328) (2009).
316. Foley, C. N. *et al.* A Fast and Efficient Colocalization Algorithm for Identifying Shared Genetic Risk Factors across Multiple Traits. *bioRxiv*. doi:[10.1101/592238](https://doi.org/10.1101/592238) (2019).
317. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genetics* **10** (ed Williams, S. M.) e1004383. doi:[10.1371/journal.pgen.1004383](https://doi.org/10.1371/journal.pgen.1004383) (2014).
318. Cervantes, J. L., Weinerman, B., Basole, C. & Salazar, J. C. TLR8: The Forgotten Relative Revindicated. *Cellular & Molecular Immunology* **9**, 434–438. doi:[10.1038/cmi.2012.38](https://doi.org/10.1038/cmi.2012.38) (2012).
319. Sullivan, A. L. *et al.* Serum Response Factor Utilizes Distinct Promoter- and Enhancer-Based Mechanisms to Regulate Cytoskeletal Gene Expression in Macrophages. *Molecular and Cellular Biology* **31**, 861–875. doi:[10.1128/MCB.00836-10](https://doi.org/10.1128/MCB.00836-10) (2011).
320. Marigorta, U. M. *et al.* Transcriptional Risk Scores Link GWAS to eQTLs and Predict Complications in Crohn’s Disease. *Nature Genetics* **49**, 1517–1521. doi:[10.1038/ng.3936](https://doi.org/10.1038/ng.3936) (2017).
321. Ongen, H., Brown, A. A., Delaneau, O., Panousis, N. I., Nica, A. C. & Dermitzakis, E. T. Estimating the Causal Tissues for Complex Traits and Diseases. *Nature Genetics*. doi:[10.1038/ng.3981](https://doi.org/10.1038/ng.3981) (October 2017).

322. Umans, B. D., Battle, A. & Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends in Genetics*. doi:[10.1016/j.tig.2020.08.009](https://doi.org/10.1016/j.tig.2020.08.009) (2020).
323. Cuomo, A. S. E. *et al.* Single-Cell RNA-Sequencing of Differentiating iPS Cells Reveals Dynamic Genetic Effects on Gene Expression. *Nature Communications* **11**. doi:[10.1038/s41467-020-14457-z](https://doi.org/10.1038/s41467-020-14457-z) (2020).
324. Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-Mapping Cellular QTLs with RASQUAL and ATAC-Seq. *Nature Genetics* **48**, 206–213. doi:[10.1038/ng.3467](https://doi.org/10.1038/ng.3467) (2016).
325. Wang, A. T. *et al.* Allele-Specific QTL Fine Mapping with PLASMA. *The American Journal of Human Genetics* **106**, 170–187. doi:[10.1016/j.ajhg.2019.12.011](https://doi.org/10.1016/j.ajhg.2019.12.011) (2020).
326. Mohammadi, P., Castel, S. E., Brown, A. A. & Lappalainen, T. Quantifying the Regulatory Effect Size of *Cis*-Acting Genetic Variation Using Allelic Fold Change. *Genome Research* **27**, 1872–1884. doi:[10.1101/gr.216747.116](https://doi.org/10.1101/gr.216747.116) (2017).
327. Wu, L., Shen, C., Seed Ahmed, M., Östenson, C.-G. & Gu, H. F. Adenylate Cyclase 3: A New Target for Anti-Obesity Drug Development: ADCY3 and Anti-Obesity. *Obesity Reviews* **17**, 907–914. doi:[10.1111/obr.12430](https://doi.org/10.1111/obr.12430) (2016).
328. Saeed, S. *et al.* Epigenetic Programming of Monocyte-to-Macrophage Differentiation and Trained Innate Immunity. *Science* **345**, 1251086–1251086. doi:[10.1126/science.1251086](https://doi.org/10.1126/science.1251086) (2014).
329. Pai, A. A., Pritchard, J. K. & Gilad, Y. The Genetic and Mechanistic Basis for Variation in Gene Regulation. *PLoS Genetics* **11** (ed Lappalainen, T.) e1004857. doi:[10.1371/journal.pgen.1004857](https://doi.org/10.1371/journal.pgen.1004857) (2015).
330. Choudhury, M. & Ramsey, S. A. Identifying Cell Type-Specific Transcription Factors by Integrating ChIP-Seq and eQTL Data-Application to Monocyte Gene Regulation. *Gene Regulation and Systems Biology* **10**, GRSB.S40768. doi:[10.4137/GRSB.S40768](https://doi.org/10.4137/GRSB.S40768) (2016).
331. Zeng, B. *et al.* Comprehensive Multiple eQTL Detection and Its Application to GWAS Interpretation. *Genetics* **212**, 905–918. doi:[10.1534/genetics.119.302091](https://doi.org/10.1534/genetics.119.302091) (2019).
332. Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-Localization with Schizophrenia GWAS. *The American Journal of Human Genetics* **102**, 1169–1184. doi:[10.1016/j.ajhg.2018.04.011](https://doi.org/10.1016/j.ajhg.2018.04.011) (2018).
333. Rotival, M. Characterising the Genetic Basis of Immune Response Variation to Identify Causal Mechanisms Underlying Disease Susceptibility. *HLA* **94**, 275–284. doi:[10.1111/tan.13598](https://doi.org/10.1111/tan.13598) (2019).
334. Langford, E., Schwertman, N. & Owens, M. Is the Property of Being Positively Correlated Transitive? *The American Statistician* **55**, 322–325 (2001).
335. Roda, G. *et al.* Crohn’s Disease. *Nature Reviews Disease Primers* **6**. doi:[10.1038/s41572-020-0156-2](https://doi.org/10.1038/s41572-020-0156-2) (2020).
336. Cotsapas, C. & Hafler, D. A. Immune-Mediated Disease Genetics: The Shared Basis of Pathogenesis. *Trends in Immunology* **34**, 22–26. doi:[10.1016/j.it.2012.09.001](https://doi.org/10.1016/j.it.2012.09.001) (2013).

337. David, T., Ling, S. F. & Barton, A. Genetics of Immune-Mediated Inflammatory Diseases. *Clinical & Experimental Immunology* **193**, 3–12. doi:[10.1111/cei.13101](https://doi.org/10.1111/cei.13101) (2018).
338. Ananthakrishnan, A. N. Epidemiology and Risk Factors for IBD. *Nature Reviews Gastroenterology & Hepatology* **12**, 205–217. doi:[10.1038/nrgastro.2015.34](https://doi.org/10.1038/nrgastro.2015.34) (2015).
339. De Souza, H. S. P. & Fiocchi, C. Immunopathogenesis of IBD: Current State of the Art. *Nature Reviews Gastroenterology & Hepatology* **13**, 13–27. doi:[10.1038/nrgastro.2015.186](https://doi.org/10.1038/nrgastro.2015.186) (2016).
340. Todd, J. A. Tackling Common Disease. *Nature* **411**, 537–539. doi:[10.1038/35079223](https://doi.org/10.1038/35079223) (2001).
341. Jostins, L. *et al.* Host–Microbe Interactions Have Shaped the Genetic Architecture of Inflammatory Bowel Disease. *Nature* **491**, 119–24. doi:[10.1038/nature11582](https://doi.org/10.1038/nature11582) (2012).
342. Liu, J. Z. *et al.* Association Analyses Identify 38 Susceptibility Loci for Inflammatory Bowel Disease and Highlight Shared Genetic Risk across Populations. *Nature Genetics* **47**, 979–986. doi:[10.1038/ng.3359](https://doi.org/10.1038/ng.3359) (2015).
343. Kaplan, G. G. The Global Burden of IBD: From 2015 to 2025. *Nature Reviews Gastroenterology & Hepatology* **12**, 720–727. doi:[10.1038/nrgastro.2015.150](https://doi.org/10.1038/nrgastro.2015.150) (2015).
344. Alatab, S. *et al.* The Global, Regional, and National Burden of Inflammatory Bowel Disease in 195 Countries and Territories, 1990–2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology & Hepatology* **5**, 17–30. doi:[10.1016/S2468-1253\(19\)30333-4](https://doi.org/10.1016/S2468-1253(19)30333-4) (2020).
345. Levin, A. D., Wildenberg, M. E. & van den Brink, G. R. Mechanism of Action of Anti-TNF Therapy in Inflammatory Bowel Disease. *Journal of Crohn's and Colitis* **10**, 989–997. doi:[10.1093/ecco-jcc/jjw053](https://doi.org/10.1093/ecco-jcc/jjw053) (2016).
346. Aggarwal, B. B. Signalling Pathways of the TNF Superfamily: A Double-Edged Sword. *Nature Reviews Immunology* **3**, 745–756. doi:[10.1038/nri1184](https://doi.org/10.1038/nri1184) (2003).
347. Kalliliias, G. D. & Ivashkiv, L. B. TNF Biology, Pathogenic Mechanisms and Emerging Therapeutic Strategies. *Nature Reviews Rheumatology* **12**, 49–62. doi:[10.1038/nrrheum.2015.169](https://doi.org/10.1038/nrrheum.2015.169) (2016).
348. Digby-Bell, J. L., Atreya, R., Monteleone, G. & Powell, N. Interrogating Host Immunity to Predict Treatment Response in Inflammatory Bowel Disease. *Nature Reviews Gastroenterology & Hepatology*. doi:[10.1038/s41575-019-0228-5](https://doi.org/10.1038/s41575-019-0228-5) (2019).
349. Adegbola, S. O., Sahnani, K., Warusavitarne, J., Hart, A. & Tozer, P. Anti-TNF Therapy in Crohn's Disease. *International Journal of Molecular Sciences* **19**, 2244. doi:[10.3390/ijms19082244](https://doi.org/10.3390/ijms19082244) (2018).
350. Lichtenstein, G. R. Comprehensive Review: Antitumor Necrosis Factor Agents in Inflammatory Bowel Disease and Factors Implicated in Treatment Response. *Therapeutic Advances in Gastroenterology* **6**, 269–293. doi:[10.1177/1756283X13479826](https://doi.org/10.1177/1756283X13479826) (2013).
351. Aitken, M., Kleinrock, M., Simorellis, A. & Nass, D. *The Global Use of Medicine in 2019 and Outlook to 2023: Forecasts and Areas to Watch* (IQVIA Institute, NC, USA, 2019).

352. Roda, G., Jharap, B., Neeraj, N. & Colombel, J.-F. Loss of Response to Anti-TNFs: Definition, Epidemiology, and Management: *Clinical and Translational Gastroenterology* **7**, e135. doi:[10.1038/ctg.2015.63](https://doi.org/10.1038/ctg.2015.63) (2016).
353. Ben-Horin, S., Kopylov, U. & Chowers, Y. Optimizing Anti-TNF Treatments in Inflammatory Bowel Disease. *Autoimmunity Reviews* **13**, 24–30. doi:[10.1016/j.autrev.2013.06.002](https://doi.org/10.1016/j.autrev.2013.06.002) (2014).
354. Flamant, M. & Roblin, X. Inflammatory Bowel Disease: Towards a Personalized Medicine. *Therapeutic Advances in Gastroenterology* **11**, 1756283X1774502. doi:[10.1177/1756283X17745029](https://doi.org/10.1177/1756283X17745029) (2018).
355. Kennedy, N. A. *et al.* Predictors of Anti-TNF Treatment Failure in Anti-TNF-Naive Patients with Active Luminal Crohn's Disease: A Prospective, Multicentre, Cohort Study. *The Lancet Gastroenterology & Hepatology* **4**, 341–353. doi:[10.1016/S2468-1253\(19\)30012-3](https://doi.org/10.1016/S2468-1253(19)30012-3) (2019).
356. Vermeire, S., Gils, A., Accossato, P., Lula, S. & Marren, A. Immunogenicity of Biologics in Inflammatory Bowel Disease. *Therapeutic Advances in Gastroenterology* **11**, 1756283X1775035. doi:[10.1177/1756283X17750355](https://doi.org/10.1177/1756283X17750355) (2018).
357. Rogler, G. Where Are We Heading to in Pharmacological IBD Therapy? *Pharmacological Research* **100**, 220–227. doi:[10.1016/j.phrs.2015.07.005](https://doi.org/10.1016/j.phrs.2015.07.005) (2015).
358. D'Haens, G. R. *et al.* The London Position Statement of the World Congress of Gastroenterology on Biological Therapy for IBD With the European Crohn's and Colitis Organization: When to Start, When to Stop, Which Drug to Choose, and How to Predict Response?: *American Journal of Gastroenterology* **106**, 199–212. doi:[10.1038/ajg.2010.392](https://doi.org/10.1038/ajg.2010.392) (2011).
359. Ding, N. S., Hart, A. & De Cruz, P. Systematic Review: Predicting and Optimising Response to Anti-TNF Therapy in Crohn's Disease - Algorithm for Practical Management. *Alimentary Pharmacology & Therapeutics* **43**, 30–51. doi:[10.1111/apt.13445](https://doi.org/10.1111/apt.13445) (2016).
360. Kopylov, U. & Seidman, E. Predicting Durable Response or Resistance to Antitumor Necrosis Factor Therapy in Inflammatory Bowel Disease. *Therapeutic Advances in Gastroenterology* **9**, 513–526. doi:[10.1177/1756283X16638833](https://doi.org/10.1177/1756283X16638833) (2016).
361. Noor, N. M., Verstockt, B., Parkes, M. & Lee, J. C. Personalised Medicine in Crohn's Disease. *The Lancet Gastroenterology & Hepatology* **5**, 80–92. doi:[10.1016/S2468-1253\(19\)30340-1](https://doi.org/10.1016/S2468-1253(19)30340-1) (2020).
362. Arijs, I. *et al.* Mucosal Gene Signatures to Predict Response to Infliximab in Patients with Ulcerative Colitis. *Gut* **58**, 1612–1619. doi:[10.1136/gut.2009.178665](https://doi.org/10.1136/gut.2009.178665) (2009).
363. Arijs, I. *et al.* Predictive Value of Epithelial Gene Expression Profiles for Response to Infliximab in Crohn's Disease. *Inflammatory Bowel Diseases* **16**, 2090–2098. doi:[10.1002/ibd.21301](https://doi.org/10.1002/ibd.21301) (2010).
364. West, N. R. *et al.* Oncostatin M Drives Intestinal Inflammation and Predicts Response to Tumor Necrosis Factor–Neutralizing Therapy in Patients with Inflammatory Bowel Disease. *Nature Medicine* **23**, 579–589. doi:[10.1038/nm.4307](https://doi.org/10.1038/nm.4307) (2017).

365. Martin, J. C. *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493–1508.e20. doi:[10.1016/j.cell.2019.08.008](https://doi.org/10.1016/j.cell.2019.08.008) (2019).
366. Gaujoux, R. *et al.* Cell-Centred Meta-Analysis Reveals Baseline Predictors of Anti-TNF α Non-Response in Biopsy and Blood of Patients with IBD. *Gut* **68**, 604–614. doi:[10.1136/gutjnl-2017-315494](https://doi.org/10.1136/gutjnl-2017-315494) (2019).
367. Verstockt, B. *et al.* Low TREM1 Expression in Whole Blood Predicts Anti-TNF Response in Inflammatory Bowel Disease. *EBioMedicine* **40**, 733–742. doi:[10.1016/j.ebiom.2019.01.027](https://doi.org/10.1016/j.ebiom.2019.01.027) (2019).
368. Verstockt, B. *et al.* TREM-1, the Ideal Predictive Biomarker for Endoscopic Healing in Anti-TNF-Treated Crohn's Disease Patients? *Gut* **68**, 1531–1533. doi:[10.1136/gutjnl-2018-316845](https://doi.org/10.1136/gutjnl-2018-316845) (2019).
369. Burke, K. E. *et al.* Genetic Markers Predict Primary Nonresponse and Durable Response to Anti-Tumor Necrosis Factor Therapy in Ulcerative Colitis. *Inflammatory Bowel Diseases* **24**, 1840–1848. doi:[10.1093/ibd/izy083](https://doi.org/10.1093/ibd/izy083) (2018).
370. Sazonovs, A. *et al.* HLA-DQA1*05 Carriage Associated With Development of Anti-Drug Antibodies to Infliximab and Adalimumab in Patients With Crohn's Disease. *Gastroenterology*. doi:[10.1053/j.gastro.2019.09.041](https://doi.org/10.1053/j.gastro.2019.09.041) (2019).
371. Simon Andrews. *FastQC: A Quality Control Tool for High Throughput Sequence Data* <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2015).
372. Dobin, A. *et al.* STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **29**, 15–21. doi:[10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635) (2013).
373. Smith, T., Heger, A. & Sudbery, I. UMI-Tools: Modeling Sequencing Errors in Unique Molecular Identifiers to Improve Quantification Accuracy. *Genome Research* **27**, 491–499. doi:[10.1101/gr.209601.116](https://doi.org/10.1101/gr.209601.116) (2017).
374. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics (Oxford, England)* **30**, 923–930. doi:[10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656) (2014).
375. Hoffman, G. E. & Schadt, E. E. variancePartition: Interpreting Drivers of Variation in Complex Gene Expression Studies. *BMC Bioinformatics* **17**. doi:[10.1186/s12859-016-1323-z](https://doi.org/10.1186/s12859-016-1323-z) (2016).
376. Cinelli, C., Forney, A. & Pearl, J. *A Crash Course in Good and Bad Controls* R-493 (2020), 10.
377. Aryee, M. J. *et al.* Minfi: A Flexible and Comprehensive Bioconductor Package for the Analysis of Infinium DNA Methylation Microarrays. *Bioinformatics* **30**, 1363–1369. doi:[10.1093/bioinformatics/btu049](https://doi.org/10.1093/bioinformatics/btu049) (2014).
378. Houseman, E. A. *et al.* DNA Methylation Arrays as Surrogate Measures of Cell Mixture Distribution. *BMC Bioinformatics* **13**, 86. doi:[10.1186/1471-2105-13-86](https://doi.org/10.1186/1471-2105-13-86) (2012).

379. MacKinnon, D. P., Krull, J. L. & Lockwood, C. M. Equivalence of the Mediation, Confounding and Suppression Effect. *Prevention science : the official journal of the Society for Prevention Research* **1**, 173 (2000).
380. Suzuki, E., Shinozaki, T. & Yamamoto, E. Causal Diagrams: Pitfalls and Tips. *Journal of Epidemiology* **30**, 153–162. doi:[10.2188/jea.JE20190192](https://doi.org/10.2188/jea.JE20190192) (2020).
381. Piasecka, B. *et al.* Distinctive Roles of Age, Sex, and Genetics in Shaping Transcriptional Variation of Human Immune Responses to Microbial Challenges. *Proceedings of the National Academy of Sciences* **115**, E488–E497. doi:[10.1073/pnas.1714765115](https://doi.org/10.1073/pnas.1714765115) (2018).
382. Hoffman, G. E. & Roussos, P. Dream: Powerful Differential Expression Analysis for Repeated Measures Designs. *Bioinformatics* (ed Gorodkin, J.) doi:[10.1093/bioinformatics/btaa687](https://doi.org/10.1093/bioinformatics/btaa687) (2020).
383. McNeish, D. Small Sample Methods for Multilevel Modeling: A Colloquial Elucidation of REML and the Kenward-Roger Correction. *Multivariate Behavioral Research* **52**, 661–670. doi:[10.1080/00273171.2017.1344538](https://doi.org/10.1080/00273171.2017.1344538) (2017).
384. Perperoglou, A., Sauerbrei, W., Abrahamowicz, M. & Schmid, M. A Review of Spline Function Procedures in R. *BMC Medical Research Methodology* **19**. doi:[10.1186/s12874-019-0666-3](https://doi.org/10.1186/s12874-019-0666-3) (2019).
385. Müllner, D. **Fastcluster** : Fast Hierarchical, Agglomerative Clustering Routines for *R* and *Python*. *Journal of Statistical Software* **53**. doi:[10.18637/jss.v053.i09](https://doi.org/10.18637/jss.v053.i09) (2013).
386. Tibshirani, R., Walther, G. & Hastie, T. Estimating the Number of Clusters in a Data Set via the Gap Statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 411–423. doi:[10.1111/1467-9868.00293](https://doi.org/10.1111/1467-9868.00293) (2001).
387. Salvador-Martín, S. *et al.* Gene Signatures of Early Response to Anti-TNF Drugs in Pediatric Inflammatory Bowel Disease. *International Journal of Molecular Sciences* **21**, 3364. doi:[10.3390/ijms21093364](https://doi.org/10.3390/ijms21093364) (2020).
388. Zalocusky, K. A. *et al.* The 10,000 Immunomes Project: Building a Resource for Human Immunology. *Cell Reports* **25**, 513–522.e3. doi:[10.1016/j.celrep.2018.09.021](https://doi.org/10.1016/j.celrep.2018.09.021) (2018).
389. Pellegrino Coppola, D. *et al.* Correction for Both Common and Rare Cell Types in Blood Is Important to Identify Genes That Correlate with Age. doi:[10.1101/2020.05.28.120600](https://doi.org/10.1101/2020.05.28.120600) (2020).
390. Boyapati, R. K., Rossi, A. G., Satsangi, J. & Ho, G.-T. Gut Mucosal DAMPs in IBD: From Mechanisms to Therapeutic Implications. *Mucosal Immunology* **9**, 567–582. doi:[10.1038/mi.2016.14](https://doi.org/10.1038/mi.2016.14) (2016).
391. Crocker, P. R., Paulson, J. C. & Varki, A. Siglecs and Their Roles in the Immune System. *Nature Reviews Immunology* **7**, 255–266. doi:[10.1038/nri2056](https://doi.org/10.1038/nri2056) (4 2007).
392. Martinez, F. O. The Transcriptome of Human Monocyte Subsets Begins to Emerge. *Journal of Biology* **8**, 99. doi:[10.1186/jbiol206](https://doi.org/10.1186/jbiol206) (2009).

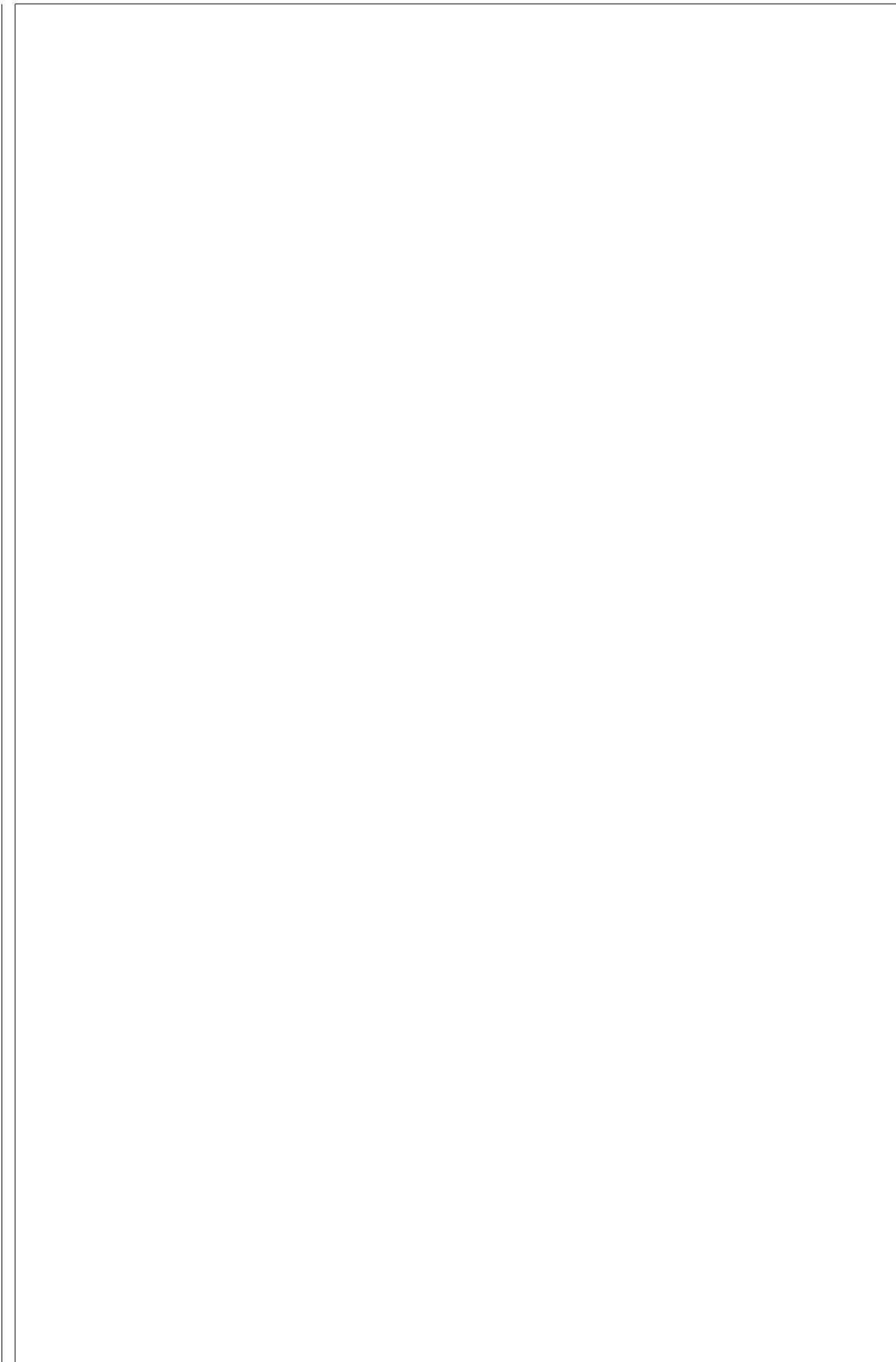
393. Villani, A.-C. *et al.* Single-Cell RNA-Seq Reveals New Types of Human Blood Dendritic Cells, Monocytes, and Progenitors. *Science* **356**, eaah4573. doi:[10.1126/science.aah4573](https://doi.org/10.1126/science.aah4573) (2017).
394. Ivashkiv, L. B. & Donlin, L. T. Regulation of Type I Interferon Responses. *Nature Reviews Immunology* **14**, 36–49. doi:[10.1038/nri3581](https://doi.org/10.1038/nri3581) (1 2014).
395. Lu, Y., Li, X., Liu, S., Zhang, Y. & Zhang, D. Toll-like Receptors and Inflammatory Bowel Disease. *Frontiers in Immunology* **9**. doi:[10.3389/fimmu.2018.00072](https://doi.org/10.3389/fimmu.2018.00072) (2018).
396. Corridoni, D., Chapman, T., Ambrose, T. & Simmons, A. Emerging Mechanisms of Innate Immunity and Their Translational Potential in Inflammatory Bowel Disease. *Frontiers in Medicine* **5**. doi:[10.3389/fmed.2018.00032](https://doi.org/10.3389/fmed.2018.00032) (2018).
397. Prame Kumar, K., Nicholls, A. J. & Wong, C. H. Y. Partners in Crime: Neutrophils and Monocytes/Macrophages in Inflammation and Disease. *Cell and Tissue Research* **371**, 551–565. doi:[10.1007/s00441-017-2753-2](https://doi.org/10.1007/s00441-017-2753-2) (2018).
398. Lügering, A., Schmidt, M., Lügering, N., Pauels, H.-G., Domschke, W. & Kucharzik, T. Infliximab Induces Apoptosis in Monocytes from Patients with Chronic Active Crohn's Disease by Using a Caspase-Dependent Pathway. *Gastroenterology* **121**, 1145–1157. doi:[10.1053/gast.2001.28702](https://doi.org/10.1053/gast.2001.28702) (2001).
399. Pararasa, C. *et al.* Reduced CD27-IgD- B Cells in Blood and Raised CD27-IgD- B Cells in Gut-Associated Lymphoid Tissue in Inflammatory Bowel Disease. *Frontiers in Immunology* **10**. doi:[10.3389/fimmu.2019.00361](https://doi.org/10.3389/fimmu.2019.00361) (2019).
400. Tretina, K., Park, E.-S., Maminska, A. & MacMicking, J. D. Interferon-Induced Guanylate-Binding Proteins: Guardians of Host Defense in Health and Disease. *Journal of Experimental Medicine* **216**, 482–500. doi:[10.1084/jem.20182031](https://doi.org/10.1084/jem.20182031) (2019).
401. Liu, S.-Y., Sanchez, D. J., Aliyari, R., Lu, S. & Cheng, G. Systematic Identification of Type I and Type II Interferon-Induced Antiviral Factors. *Proceedings of the National Academy of Sciences* **109**, 4239–4244. doi:[10.1073/pnas.1114981109](https://doi.org/10.1073/pnas.1114981109) (2012).
402. Van Baarsen, L. G. *et al.* Regulation of IFN Response Gene Activity during Infliximab Treatment in Rheumatoid Arthritis Is Associated with Clinical Response to Treatment. *Arthritis Research & Therapy* **12**, R11. doi:[10.1186/ar2912](https://doi.org/10.1186/ar2912) (2010).
403. Laukoetter, M. *et al.* O-014: IFN-Gamma Induces Apoptosis in Inflammation by Inhibition of the Wnt-Pathway. *Inflammatory Bowel Diseases* **14**, S4–S5. doi:[10.1097/00054725-20081001-00014](https://doi.org/10.1097/00054725-20081001-00014) (suppl_1 2008).
404. Mavragani, C. P., La, D. T., Stohl, W. & Crow, M. K. Association of the Response to Tumor Necrosis Factor Antagonists with Plasma Type I Interferon Activity and Interferon- β/α Ratios in Rheumatoid Arthritis Patients: A Post Hoc Analysis of a Predominantly Hispanic Cohort. *Arthritis & Rheumatism* **62**, 392–401. doi:[10.1002/art.27226](https://doi.org/10.1002/art.27226) (2010).
405. Wright, H. L., Thomas, H. B., Moots, R. J. & Edwards, S. W. Interferon Gene Expression Signature in Rheumatoid Arthritis Neutrophils Correlates with a Good Response to TNFi Therapy. *Rheumatology* **54**, 188–193. doi:[10.1093/rheumatology/keu299](https://doi.org/10.1093/rheumatology/keu299) (2015).

406. Gutierrez-Arcelus, M. *et al.* Allele-Specific Expression Changes Dynamically during T Cell Activation in HLA and Other Autoimmune Loci. *Nature Genetics* **52**, 247–253. doi:[10.1038/s41588-020-0579-4](https://doi.org/10.1038/s41588-020-0579-4) (2020).
407. Tracey, D., Klareskog, L., Sasso, E. H., Salfeld, J. G. & Tak, P. P. Tumor Necrosis Factor Antagonist Mechanisms of Action: A Comprehensive Review. *Pharmacology & Therapeutics* **117**, 244–279. doi:[10.1016/j.pharmthera.2007.10.001](https://doi.org/10.1016/j.pharmthera.2007.10.001) (2008).
408. Gibson, D. J. *et al.* Review Article: Determination of the Therapeutic Range for Therapeutic Drug Monitoring of Adalimumab and Infliximab in Patients with Inflammatory Bowel Disease. *Alimentary Pharmacology & Therapeutics* **51**, 612–628. doi:[10.1111/apt.15643](https://doi.org/10.1111/apt.15643) (2020).
409. Liu, X. *Methods for Handling Missing Data in Methods and Applications of Longitudinal Data Analysis* 441–473 (Elsevier, 2016). doi:[10.1016/B978-0-12-801342-7.00014-9](https://doi.org/10.1016/B978-0-12-801342-7.00014-9).
410. Ibrahim, J. G. & Molenberghs, G. Missing Data Methods in Longitudinal Studies: A Review. *TEST* **18**, 1–43. doi:[10.1007/s11749-009-0138-x](https://doi.org/10.1007/s11749-009-0138-x) (2009).
411. Imhann, F. *et al.* The 1000IBD Project: Multi-Omics Data of 1000 Inflammatory Bowel Disease Patients; Data Release 1. *BMC Gastroenterology* **19**. doi:[10.1186/s12876-018-917-5](https://doi.org/10.1186/s12876-018-917-5) (2019).
412. Toonen, E. J. M. *et al.* Validation Study of Existing Gene Expression Signatures for Anti-TNF Treatment in Patients with Rheumatoid Arthritis. *PLOS ONE* **7**, e33199. doi:[10.1371/journal.pone.0033199](https://doi.org/10.1371/journal.pone.0033199) (2012).
413. Li, S. *et al.* Metabolic Phenotypes of Response to Vaccination in Humans. *Cell* **169**, 862–877.e17. doi:[10.1016/j.cell.2017.04.026](https://doi.org/10.1016/j.cell.2017.04.026) (2017).
414. Breuer, K. *et al.* InnateDB: Systems Biology of Innate Immunity and beyond—Recent Updates and Continuing Curation. *Nucleic Acids Research* **41**, D1228–D1233. doi:[10.1093/nar/gks1147](https://doi.org/10.1093/nar/gks1147) (2013).
415. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P. & Mesirov, J. P. Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740. doi:[10.1093/bioinformatics/btr260](https://doi.org/10.1093/bioinformatics/btr260) (2011).
416. Senn, S. Statistical Pitfalls of Personalized Medicine. *Nature* **563**, 619–621. doi:[10.1038/d41586-018-07535-2](https://doi.org/10.1038/d41586-018-07535-2) (2018).
417. Senn, S. Mastering Variation: Variance Components and Personalised Medicine. *Statistics in Medicine* **35**, 966–977. doi:[10.1002/sim.6739](https://doi.org/10.1002/sim.6739) (2016).
418. Farahbod, M. & Pavlidis, P. Untangling the Effects of Cellular Composition on Coexpression Analysis. *Genome Research*, gr.256735.119. doi:[10.1101/gr.256735.119](https://doi.org/10.1101/gr.256735.119) (2020).
419. Leek, J. T. Svaseq: Removing Batch Effects and Other Unwanted Noise from Sequencing Data. *Nucleic Acids Research* **42**, e161–e161. doi:[10.1093/nar/gku864](https://doi.org/10.1093/nar/gku864) (2014).
420. Liu, Q. & Markatou, M. Evaluation of Methods in Removing Batch Effects on RNA-Seq Data. *Infectious Diseases and Translational Medicine* **2**, 3–9. doi:[10.11979/idtm.201601002](https://doi.org/10.11979/idtm.201601002) (2016).

421. Van der Wijst, M. *et al.* The Single-Cell eQTLGen Consortium. *eLife* **9**. doi:[10.7554/eLife.52155](https://doi.org/10.7554/eLife.52155) (2020).
422. Efron, B. Prediction, Estimation, and Attribution. *Journal of the American Statistical Association* **115**, 636–655. doi:[10.1080/01621459.2020.1762613](https://doi.org/10.1080/01621459.2020.1762613) (2020).
423. Kazmin, D. *et al.* Systems Analysis of Protective Immune Responses to RTS,S Malaria Vaccination in Humans. *Proceedings of the National Academy of Sciences* **114**, 2425–2430. doi:[10.1073/pnas.1621489114](https://doi.org/10.1073/pnas.1621489114) (2017).
424. Haynes, B. F. *et al.* Immune-Correlates Analysis of an HIV-1 Vaccine Efficacy Trial. *New England Journal of Medicine* **366** (ed Kremer, E. J.) 1275–1286. doi:[10.1056/NEJMoa1113425](https://doi.org/10.1056/NEJMoa1113425) (2012).
425. Zak, D. E. *et al.* Merck Ad5/HIV Induces Broad Innate Immune Activation That Predicts CD8+ T-Cell Responses but Is Attenuated by Preexisting Ad5 Immunity. *Proceedings of the National Academy of Sciences* **109**, E3503–E3512. doi:[10.1073/pnas.1208972109](https://doi.org/10.1073/pnas.1208972109) (2012).
426. Fourati, S. *et al.* Pre-Vaccination Inflammation and B-Cell Signalling Predict Age-Related Hyporesponse to Hepatitis B Vaccination. *Nature Communications* **7**, 10369. doi:[10.1038/ncomms10369](https://doi.org/10.1038/ncomms10369) (2016).
427. Qi, Q. *et al.* Defective T Memory Cell Differentiation after Varicella Zoster Vaccination in Older Individuals. *PLOS Pathogens* **12** (ed Rooney, C. M.) e1005892. doi:[10.1371/journal.ppat.1005892](https://doi.org/10.1371/journal.ppat.1005892) (2016).
428. Rechtien, A. *et al.* Systems Vaccinology Identifies an Early Innate Immune Signature as a Correlate of Antibody Responses to the Ebola Vaccine rVSV-ZEBOV. *Cell Reports* **20**, 2251–2261. doi:[10.1016/j.celrep.2017.08.023](https://doi.org/10.1016/j.celrep.2017.08.023) (2017).
429. Avey, S. *et al.* Multiple Network-Constrained Regressions Expand Insights into Influenza Vaccination Responses. *Bioinformatics* **33**, i208–i216. doi:[10.1093/bioinformatics/btx260](https://doi.org/10.1093/bioinformatics/btx260) (2017).
430. Avey, S. *et al.* Seasonal Variability and Shared Molecular Signatures of Inactivated Influenza Vaccination in Young and Older Adults. *The Journal of Immunology* **204**, 1661–1673. doi:[10.4049/jimmunol.1900922](https://doi.org/10.4049/jimmunol.1900922) (2020).
431. MAQC Consortium. The MicroArray Quality Control (MAQC)-II Study of Common Practices for the Development and Validation of Microarray-Based Predictive Models. *Nature Biotechnology* **28**, 827–838. doi:[10.1038/nbt.1665](https://doi.org/10.1038/nbt.1665) (2010).
432. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-Generation Machine Learning for Biological Networks. *Cell* **173**, 1581–1592. doi:[10.1016/j.cell.2018.05.015](https://doi.org/10.1016/j.cell.2018.05.015) (2018).
433. Chibon, F. Cancer Gene Expression Signatures – The Rise and Fall? *European Journal of Cancer* **49**, 2000–2009. doi:[10.1016/j.ejca.2013.02.021](https://doi.org/10.1016/j.ejca.2013.02.021) (2013).

434. Michiels, S., Ternès, N. & Rotolo, F. Statistical Controversies in Clinical Research: Prognostic Gene Signatures Are Not (yet) Useful in Clinical Practice. *Annals of Oncology* **27**, 2160–2167. doi:[10.1093/annonc/mdw307](https://doi.org/10.1093/annonc/mdw307) (2016).
435. Kwa, M., Makris, A. & Esteva, F. J. Clinical Utility of Gene-Expression Signatures in Early Stage Breast Cancer. *Nature Reviews Clinical Oncology* **14**, 595–610. doi:[10.1038/nrclinonc.2017.74](https://doi.org/10.1038/nrclinonc.2017.74) (2017).
436. Davey Smith, G. & Hemani, G. Mendelian Randomization: Genetic Anchors for Causal Inference in Epidemiological Studies. *Human Molecular Genetics* **23**, R89–R98. doi:[10.1093/hmg/ddu328](https://doi.org/10.1093/hmg/ddu328) (2014).
437. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the Potential Role of Pleiotropy in Mendelian Randomization Studies. *Human Molecular Genetics* **27**, R195–R208. doi:[10.1093/hmg/ddy163](https://doi.org/10.1093/hmg/ddy163) (2018).
438. Neumeyer, S., Hemani, G. & Zeggini, E. Strengthening Causal Inference for Complex Disease Using Molecular Quantitative Trait Loci. *Trends in Molecular Medicine* **26**, 232–241. doi:[10.1016/j.molmed.2019.10.004](https://doi.org/10.1016/j.molmed.2019.10.004) (2020).
439. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian Randomisation Studies: A Guide, Glossary, and Checklist for Clinicians. *BMJ*, k601. doi:[10.1136/bmj.k601](https://doi.org/10.1136/bmj.k601) (2018).
440. Davey Smith, G., Holmes, M. V., Davies, N. M. & Ebrahim, S. Mendel's Laws, Mendelian Randomization and Causal Inference in Observational Data: Substantive and Nomenclatural Issues. *European Journal of Epidemiology* **35**, 99–111. doi:[10.1007/s10654-020-00622-7](https://doi.org/10.1007/s10654-020-00622-7) (2020).
441. Gusev, A. *et al.* Integrative Approaches for Large-Scale Transcriptome-Wide Association Studies. *Nature Genetics* **48**, 245–252. doi:[10.1038/ng.3506](https://doi.org/10.1038/ng.3506) (2016).
442. Zhu, H. & Zhou, X. Transcriptome-Wide Association Studies: A View from Mendelian Randomization. *Quantitative Biology*. doi:[10.1007/s40484-020-0207-4](https://doi.org/10.1007/s40484-020-0207-4) (2020).
443. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the Causal Relationship between Imprecisely Measured Traits Using GWAS Summary Data. *PLOS Genetics* **13** (ed Li, J.) e1007081. doi:[10.1371/journal.pgen.1007081](https://doi.org/10.1371/journal.pgen.1007081) (2017).
444. Wang, L. & Michoel, T. Efficient and Accurate Causal Inference with Hidden Confounders from Genome-Transcriptome Variation Data. *PLOS Computational Biology* **13** (ed Listgarten, J.) e1005703. doi:[10.1371/journal.pcbi.1005703](https://doi.org/10.1371/journal.pcbi.1005703) (2017).
445. Munafò, M. R. & Davey Smith, G. Robust Research Needs Many Lines of Evidence. *Nature* **553**, 399–401. doi:[10.1038/d41586-018-01023-3](https://doi.org/10.1038/d41586-018-01023-3) (2018).
446. Taylor, D. L. *et al.* Integrative Analysis of Gene Expression, DNA Methylation, Physiological Traits, and Genetic Variation in Human Skeletal Muscle. *Proceedings of the National Academy of Sciences* **116**, 10883–10888. doi:[10.1073/pnas.1814263116](https://doi.org/10.1073/pnas.1814263116) (2019).

447. Zheng, J. *et al.* Phenome-Wide Mendelian Randomization Mapping the Influence of the Plasma Proteome on Complex Diseases. *Nature Genetics* **52**, 1122–1131. doi:[10.1038/s41588-020-0682-6](https://doi.org/10.1038/s41588-020-0682-6) (2020).



List of Abbreviations

AC allele count

APC antigen-presenting cell

ASC antibody-secreting cell

ASE allele-specific expression

AUC area under the curve

BCR B cell receptor

BH Benjamini-Hochberg

BMI body mass index

BTM blood transcription module

CD Crohn's disease

CDR complementarity-determining region

ChIP-seq chromatin immunoprecipitation sequencing

CPM counts per million

CRP C-reactive protein

CyTOF cytometry by time-of-flight

DAMP damage-associated molecular pattern

DC dendritic cell

df degree of freedom

DGE differential gene expression

ELISA enzyme-linked immunosorbent assay

ELISPOT enzyme-linked immune absorbent spot

eQTL expression quantitative trait locus
FACS fluorescence-activated cell sorting
FC fold change
FDR false discovery rate
FWER family-wise error rate
GO Gene Ontology
GWAS genome-wide association study
HA haemagglutinin
HAI haemagglutination inhibition
HBI Harvey Bradshaw index
HIRD Human Immune Response Dynamics
HLA human leukocyte antigen
HSC hematopoietic stem cell
HWE Hardy-Weinberg equilibrium
IBD inflammatory bowel disease
IIV inactivated influenza vaccine
IMID immune-mediated inflammatory disease
INT inverse normal transformation
IV instrumental variable
LAIIV live attenuated influenza vaccine
LD linkage disequilibrium
LFSR local false sign rate
LMM linear mixed model
LOCO leave-one-chromosome-out
LOR loss of response
LRT likelihood ratio test
MAF minor allele frequency

MANOVA	multivariate analysis of variance
MAR	missing at random
MCAR	missing completely at random
MHC	major histocompatibility complex
ML	maximum likelihood
MN	microneutralisation
MNAR	missing not at random
molQTL	molecular quantitative trait locus
MR	Mendelian randomisation
mRNA	messenger RNA
MS	multiple sclerosis
NA	neuraminidase
ncRNA	non-coding RNA
NK	natural killer
PAMP	pathogen-associated molecular pattern
PANTS	Personalised Anti-TNF Therapy in Crohn's Disease
PBMC	peripheral blood mononuclear cell
PC	principal component
PCA	principal component analysis
PNR	primary non-response
PRR	pattern recognition receptor
PVE	proportion of variance explained
QTL	quantitative trait locus
RA	rheumatoid arthritis
RBC	red blood cell
REML	restricted maximum likelihood
reQTL	response expression quantitative trait locus

RNA-seq	RNA sequencing
rRNA	ribosomal RNA
scRNA-seq	single-cell RNA sequencing
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
T1D	type 1 diabetes
T2D	type 2 diabetes
TF	transcription factor
TIV	trivalent inactivated influenza vaccine
TLR	toll-like receptor
TMM	trimmed mean of M-values
TNF	tumour necrosis factor
TPM	transcripts per million
TRI	titre response index
TSS	transcription start site
TWAS	transcriptome-wide association study
UC	ulcerative colitis
UTR	untranslated region
WES	whole-exome sequencing
WGS	whole-genome sequencing
WHO	World Health Organization