

<title>

Benjamin Yu Hang Bai

2020-04-10 04:27:50+01:00

<dedication>

Abstract

<thesis abstract>

Acknowledgements

<acknowledgements>

Contents

List of Figures	ix
List of Tables	xiii
1 Transcriptomic response to influenza A (H1N1)pdm09 vaccine	1
1.1 Introduction	1
1.1.1 Seasonal and pandemic influenza	1
1.1.2 Quantifying immune response to influenza vaccines . .	2
1.1.3 Systems vaccinology of influenza vaccines	2
1.1.4 The Human Immune Response Dynamics (HIRD) study	3
1.1.5 Chapter summary	4
1.2 Methods	4
1.2.1 Existing HIRD study data and additional data	4
1.2.2 Computing baseline-adjusted measures of antibody response	5
1.2.3 Genotype data generation	6
1.2.4 Genotype data preprocessing	6
1.2.5 Computing genotype principal components as covariates for ancestry	10
1.2.6 Genotype phasing and imputation	10
1.2.7 RNA-seq data generation	10
1.2.8 RNA-seq quantification and filtering	12
1.2.9 Array data preprocessing	14
1.2.10 Differential gene expression	17
1.2.10.1 Per-platform differential gene expression model	20

1.2.10.2	Choice of differential gene expression meta-analysis method	20
1.2.10.3	Prior for between-studies heterogeneity	21
1.2.10.4	Prior for effect size	22
1.2.10.5	Evaluation of priors	22
1.2.10.6	Multiple testing correction	22
1.2.11	Gene set enrichment analysis using blood transcription modules	24
1.3	Results	24
1.3.1	Extensive global changes in expression after vaccination	24
1.3.2	Innate immune response at day 1 post-vaccination .	24
1.3.3	Adaptive immune response at day 7 post-vaccination .	26
1.3.4	Expression signatures associated with antibody response	26
1.3.5	Identifying expression signatures for predicting antibody response [probably cut this section and just add to discussion]	28
1.4	Discussion	28
A	Supplementary Materials	35
A.1	Chapter 2	35
A.2	Chapter 3	35
A.3	Chapter 4	36
Bibliography		37
List of Abbreviations		45

List of Figures

1.1	Data types, timepoints, and sample sizes. Individuals were vaccinated after day 0 sampling. Antibodies to the vaccine strain were measured by haemagglutination inhibition (HAI) and microneutralisation (MN) assays. Array and RNA-sequencing (RNA-seq) gene expression measured in the peripheral blood mononuclear cell (PBMC) compartment.	5
1.2	Comparison of titre response index (TRI) to HAI (left column) and MN (right column) titres and binary responder/non-responder status (colored) in 166 Human Immune Response Dynamics (HIRD) individuals. Row 1: baseline titres are positively correlated to post-vaccination titres. Row 2: baseline titres are negatively correlated to fold change. Row 3: TRI regresses out the correlation between baseline titre and response. Row 4: TRI is still comparable in ordering to binary response status.	7
1.3	Distribution of TRI, stratified by platform used to measure expression.	8
1.4	Sample filters for missingness and heterozygosity rate. Samples outside the central rectangle were excluded.	9
1.5	HIRD samples (cyan) projected onto principal component (PC)1 and PC2 axes defined by principal component analysis (PCA) of HapMap 3 samples. The first two PCs separate European (CEU, upper-right) from Asian (CHB and JPT, lower-right) and African (YRI, lower-left) individuals.	11
1.6	FastQC sequence quality versus read position for HIRD RNA-seq samples.	12
1.7	FastQC sequence duplication levels for HIRD RNA-seq samples.	13

1.8	FastQC GC profile for HIRD RNA-seq samples.	13
1.9	Distributions of removed short ncRNA and globin counts as a proportion of total counts in RNA-seq samples.	15
1.10	Distribution of the proportion of samples in which genes were detected (non-zero expression). Many genes are not detected in any samples. Vertical line shows 5% threshold below which genes were discarded.	15
1.11	Distribution of gene expressions for RNA-seq samples before and after filtering no expression and low expression genes. Vertical line shown at counts per million (CPM) = 0.5 threshold.	16
1.12	Raw foreground intensities for 173 HIRD array samples. Colored by array processing batch.	16
1.13	Array intensity estimates after VSN normalisation and collapsing of probes to genes. Colored by array processing batch.	18
1.14	First four PCs in the HIRD expression data, colored by platform and batch (left), and timepoint (right).	19
1.15	Gamma prior for τ used for bayesmeta (blue), compared to the empirical distribution of per-gene frequentist metafor::rma estimates for τ , for the day 1 vs. baseline effect (small estimates of $\tau < 0.01$ excluded). Empirical log-normal fit also shown (red).	23
1.16	Normal prior for μ used for bayesmeta (blue), compared to the empirical distribution of per-gene frequentist metafor::rma estimates for τ , for the day 1 vs. baseline effect. The non-scaled normal fit is shown (black), as well as a Cauchy fit (red). . .	23
1.17	Normalised gene expression for genes differentially expressed between any pair of timepoints ($lfsr < 0.05$, absolute fold change > 1.5) across HIRD samples, clustered by gene (Manhattan distance metric).	25
1.18	Transcriptomic modules significantly up or downregulated post-vaccination. Size of circle indicates effect size. Color of circle indicates significance and direction of effect (red = upregulation, blue = downregulation).	27
1.19	DGE effect sizes estimated in array vs. RNA-seq. Significance colored by frequentist random effects meta-analysis FDR < 0.05 . Genes with day 7 expression associated with responder/non-responder status in ¹⁸ are circled for that contrast.	29

LIST OF FIGURES

LIST OF FIGURES

List of Tables

1.1	Sample descriptive statistics	33
1.2	HIRD batch balance	34

Chapter 1

Transcriptomic response to influenza A (H1N1)pdm09 vaccine

1.1 Introduction

1.1.1 Seasonal and pandemic influenza

Influenza is an infectious disease, generally seasonal, caused by the influenza A and influenza B viruses in humans. Influenza A viruses circulate not only in humans, but also in a variety of other birds and mammal hosts. They are classified into antigenically-distinct subtypes by the combination of two surface proteins: haemagglutinin (HA) and neuraminidase (NA)¹.

There are three classes of influenza vaccine against seasonal strains in use: inactivated vaccines, live attenuated influenza vaccines (LAIVs), and recombinant HA vaccines. These vaccines confer a degree of strain-specific protection, primarily by raising serum antibodies against the HA and/or NA proteins. Antigenic drift, the accumulation of mutations in these surface proteins over time, necessitates the annual reformulation of seasonal influenza vaccines to reflect circulating strains^{2,3}. On occasion, a novel subtype against which the majority of the population is immunologically naive can arise, often from zoonotic origins. A recent example occurred in 2009, when an outbreak of a novel swine-origin strain, eventually termed influenza A (H1N1)pdm09, resulted in a global pandemic, the fourth to occur in the last 100 years¹.

why? for diff groups of people

add a point that 2009h1n1 is now circulating seasonally

1.1.2 Quantifying immune response to influenza vaccines

Add specific section about pandemrix? or maybe in methods

The 2009 pandemic motivated the rapid development, trialing, and licensing of several novel vaccines⁴. Immune response to influenza vaccines in clinical trials is evaluated by assays that measure levels of antibodies specific to the vaccine strain(s). The haemagglutination inhibition (HAI) assay measures the levels of serum antibodies specific to the HA surface protein. The related microneutralisation (MN) assay measures levels of antibodies (which may or may not be anti-HA) that neutralise the infectivity of the virus in cell culture⁵. Values from these assays can be compared against thresholds for known correlates of protection: markers that associate with whether an individual is protected from the disease. For example, HAI titres are regarded as the primary correlate of protection for inactivated influenza vaccines. Targets that regulatory agencies expect a licensed vaccine to meet are based on thresholds such as the proportion of trial individuals achieving HAI titres ≥ 40 and seroconversion (≥ 4 -fold increase in titres)^{6,7}.

1.1.3 Systems vaccinology of influenza vaccines

is there a more recent review?

Although HAI titres are accepted as established correlates for inactivated seasonal influenza vaccines, they fail to account for alternate mechanisms such as T cell-mediated protection, and correlates for LAIV and pandemic influenza vaccines are less reliable². For novel and emerging diseases, there may be no prior knowledge of robust correlates to use in the vaccine development process. In response, the last decade has seen the rise of systems vaccinology studies: the analysis of high-dimensional data measured using multiple technologies in vaccinated individuals, in order to characterise response to vaccination at multiple levels of the biological system⁸. Such information helps elucidate a vaccine's mode of action, discover "molecular signatures" predictive of vaccine safety and efficacy, and has become an increasingly important part of the modern vaccine development chain^{9,10}.

Various systems vaccinology studies of seasonal influenza vaccines have been conducted, taking longitudinal measurements pre-vaccination, and commonly at some subset of days 1, 3, 7, and 28 post-vaccination. These measurements can be correlated to changes in antibody titres after vaccination to define signatures of antibody response with potential utility as correlates of protection. One of the earliest such studies by Zhu et al.¹¹

found that expression of type 1 interferon-modulated genes was a signature of response to **LAIV**. An expression signature including *STAT1*, CD74, and E2F2 correlated with serum antibody titres after vaccination with trivalent inactivated influenza vaccine¹²; kinase CaMKIV expression is also a strong predictor¹³, as are genes related to B cell proliferation¹⁴.

For these studies of seasonal influenza vaccines in adults, responses tend to be biased by recall from past vaccination or infection^{12,15}. There have also been few studies of adjuvanted influenza vaccines, despite their superior efficacy in comparison to non-adjuvanted counterparts^{16,17}.

1.1.4 The Human Immune Response Dynamics (HIRD) study

The **Human Immune Response Dynamics (HIRD)** study conducted by Sobolev *et al.* [18] was conceived with the above limitations in mind. The vaccine studied was Pandemrix, an AS03-adjuvanted, split-virion, inactivated vaccine against the influenza A (H1N1)pdm09 strain, for which the majority of the cohort at the time would be unlikely to have immunological memory. A total of 178 individuals were vaccinated with a single dose of Pandemrix, and longitudinal transcriptomic, cellular, antibody titre, and adverse event phenotypes were collected. Gene expression was profiled using a microarray, and **differential gene expression (DGE)** analyses detected genes associated with both myeloid and lymphoid effector functions upregulated at day 1, most prominently for genes associated with interferon responses. These early myeloid responses were consistent with studies of unadjuvanted seasonal influenza vaccines, but the interferon gamma-associated lymphoid response was unique to this adjuvanted vaccine.

Genes related to plasma cell development and antibody production were more highly expressed in 23 vaccine responders compared to 18 non-responders at day 7 post-vaccination. However, due to high variability among the vaccine non-responders in variables such as baseline antibody titres, a consensus predictive model that segregated the two groups could not be built, even considering other measures such as frequencies of immune cell subsets and serum cytokine levels, suggesting there was no single contributing factor that led to vaccine failure. This is in contrast to several studies of seasonal influenza vaccines, where certain expression signatures are able to predict vaccine response even pre-vaccination^{19–22}.

1.1.5 Chapter summary

Transcriptomic measurements in the original **HIRD** study were restricted to a relatively small number (46/178) of individuals, potentially limiting power to detect expression signatures associated with antibody response. In addition, the responder vs. non-responder phenotype definition used does not account for variation in pre-existing baseline titres, and the binary definition can result in loss of statistical power^{23,24}.

In this chapter, I integrate the original microarray data from **HIRD** with **RNA-sequencing (RNA-seq)** data on a larger subset (75) of newly sequenced individuals from the same cohort using Bayesian random-effects meta-analysis. The overall pattern of expression over time from my meta-analysis agrees with the patterns from the original study¹⁸, with transient innate immune response at day 1 post-vaccination, progressing to adaptive immune response by day 7.

needs 1 more punchline sentence here

From existing **HAI** and **MN** data, I compute a baseline-adjusted, continuous measure of antibody response to vaccination, the **titre response index (TRI)**¹². Effect sizes of genes with expression that correlated with **TRI** were very dependent on measurement platform (array or **RNA-seq**), and no robust hits were detected in the meta-analysis. Leveraging the greater power that rank-based gene set enrichment analyses affords, I find modules of coexpressed genes that correlate with antibody response, with the strongest effects observed for adaptive immune modules at day 7, but also in inflammatory modules at baseline.

1.2 Methods

1.2.1 Existing HIRD study data and additional data

The design of the **HIRD** study is described in¹⁸. In brief, the study enrolled 178 healthy adult volunteers in the UK. The vaccine dose was administered after blood sampling on day 0; five other longitudinal blood samples were taken on days -7, 0, 1, 7, 14 and 63. Serological responses were measured on days -7 and 63 using the **HAI** and **MN** assays, and various subsets of the cohort were also profiled for serum cytokine levels (Luminex panel, days -7, 0, 1 and 7), immune cell subset counts (**fluorescence-activated cell sorting (FACS)** panels, all days), and **peripheral blood mononuclear cell (PBMC)**

gene expression (microarray, days -7, 0, 1 and 7).

In addition to the existing data, array genotypes were generated for 169 individuals; and **RNA-seq** data for 75 individuals at days 0, 1, and 7. The sets of individuals with gene expression assayed by microarray and **RNA-seq** is disjoint, as no biological material for RNA extraction remained for the microarray individuals. An overview of datasets is shown in Fig. 1.1.

1.2.2 Computing baseline-adjusted measures of antibody response

In¹⁸, Pandemrix responders were defined as individuals with ≥ 4 -fold titre increases in either the **HAI** or **MN** assays. This is a threshold for seroconversion set out by the U.S. Food and Drug Administration²⁵, and is used in many studies of seasonal influenza vaccines⁹. The responder status for 166 individuals with both **HAI** and **MN** titres available at baseline (day -7) and post-vaccination (day 63) were computed according to this definition. However,¹⁸ noted there was heterogeneity in the baseline titres of non-responders, citing “glass ceiling” non-responders whose high baseline titres made the fixed 4-fold threshold hard to achieve. Dichotomisation of continuous response variables can also result in loss of statistical power^{23,24}.

atm I'm not using R/NR.
wording here implies I am

cite appropriate subfig-
ures here

To address these concerns, I computed the **TRI** as defined in Bucasas *et al.* [12]. 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 standardized to zero mean and unit variance, then averaged. The **TRI** expresses a continuous

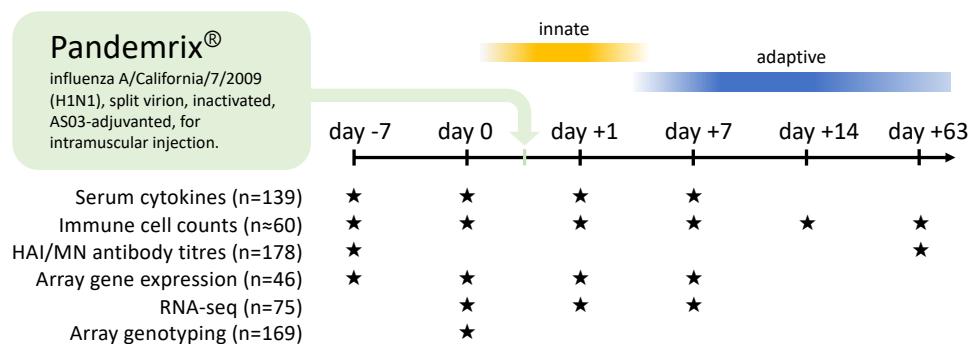


Figure 1.1: Data types, timepoints, and sample sizes. Individuals were vaccinated after day 0 sampling. Antibodies to the vaccine strain were measured by **HAI** and **MN** assays. Array and **RNA-seq** gene expression measured in the **PBMC** compartment.

measure of change in antibody titres across both assays post-vaccination, compared to individuals with a similar baseline titre, and remains comparable to the binary 4-fold change definition ([Fig. 1.2](#)).

cite appropriate subfigures here, after adding proper subfigure labels

Descriptive statistics for the 114 individuals with both gene expression and antibody titre data are presented in [Table 1.1](#). Although the proportion of responders between array (32/44) and [RNA-seq](#) (59/70) individuals is similar ($p = 0.1551$, Fisher's exact test), the variance of [TRI](#) in array individuals is higher ($p = 0.0002098$, Levene's test), suggesting more extreme antibody response phenotypes are present ([Fig. 1.3](#)). The cause of this is unknown, there is a possibility that individuals with more extreme phenotypes were prioritised for array transcriptomics in the original [HIRD](#) study*.

Add to collab note that extractions were done at KCL

1.2.3 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 (Illumina). In total, 192 samples from 176 individuals in the HIRD cohort were genotyped at 550601 markers, including replicate samples submitted for individuals where extracted DNA concentrations were low.

1.2.4 Genotype data preprocessing

Using PLINK (v1.90b3w), genotype data underwent the following quality control procedures to remove poorly genotyped samples and markers: max marker missingness across samples < 5%, max sample missingness across markers < 1%, max marker heterozygosity rate within 3 standard deviations of the mean (threshold selected visually to exclude outliers, [Fig. 1.4](#)), removal of markers that deviate from Hardy–Weinberg equilibrium (`--hwe` option, $p < 0.00001$).

To exclude highly-related individuals and deduplicate replicate samples, pairwise kinship coefficients were computed on [minor allele frequency \(MAF\)](#) < 0.05 pruned genotypes using KING (v1.4). For each pair of samples with pairwise kinship coefficient > 0.177 (first-degree relatives or closer), the sample with lower marker missingness was selected.

After filtering, 169 samples and 549414 markers remained.

*Personal communication with authors.

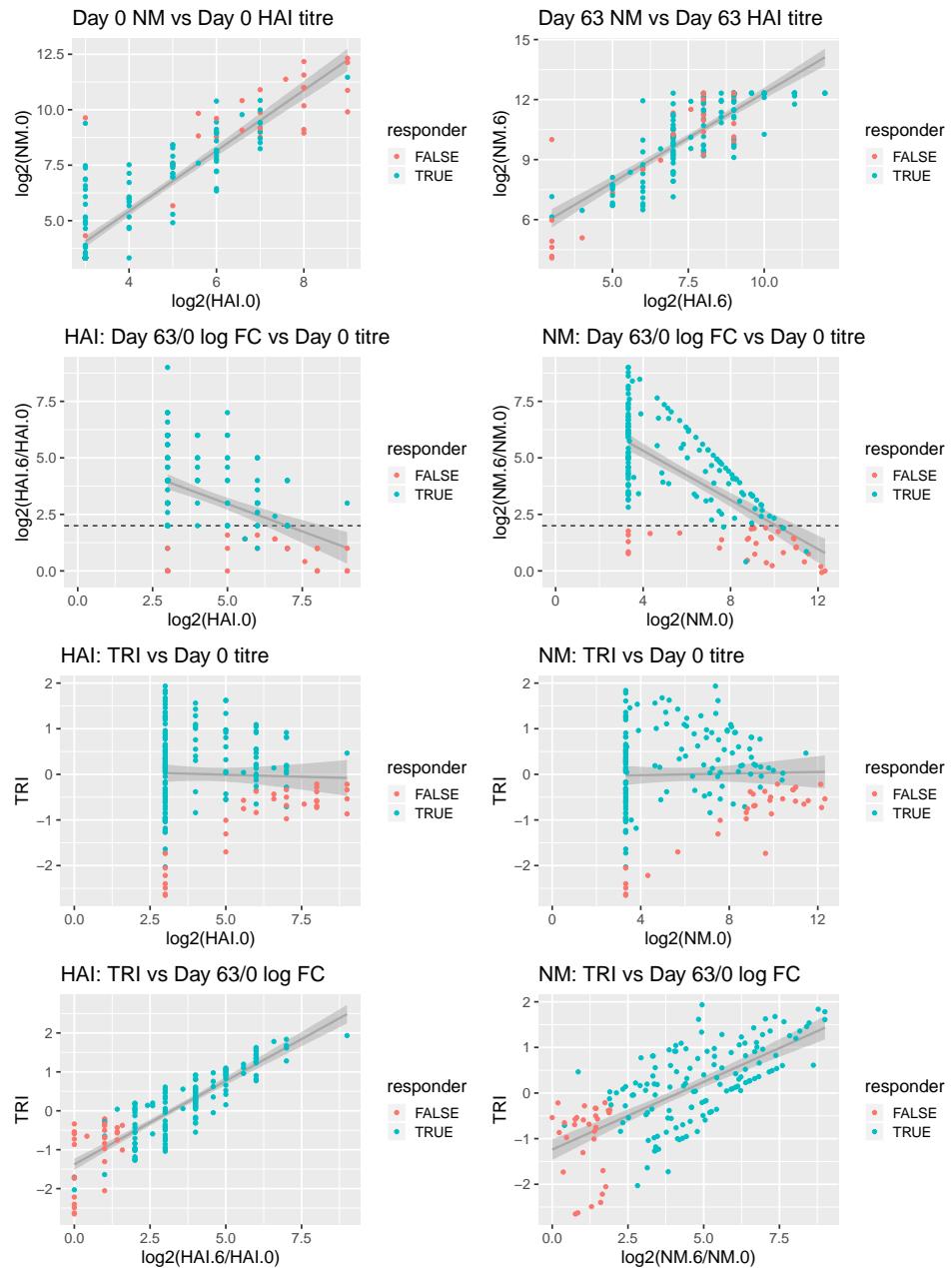


Figure 1.2: Comparison of **TRI** to **HAI** (left column) and **MN** (right column) titres and binary responder/non-responder status (colored) in 166 **HIRD** individuals. Row 1: baseline titres are positively correlated to post-vaccination titres. Row 2: baseline titres are negatively correlated to fold change. Row 3: **TRI** regresses out the correlation between baseline titre and response. Row 4: **TRI** is still comparable in ordering to binary response status.

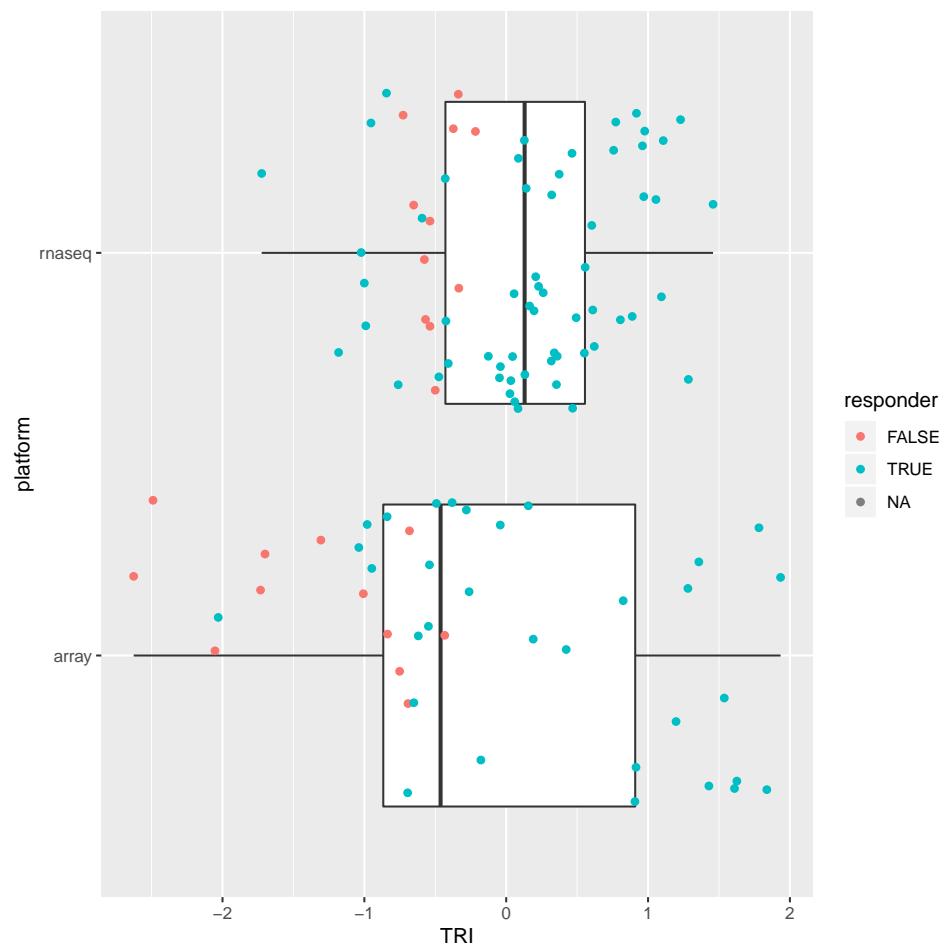


Figure 1.3: Distribution of TRI, stratified by platform used to measure expression.

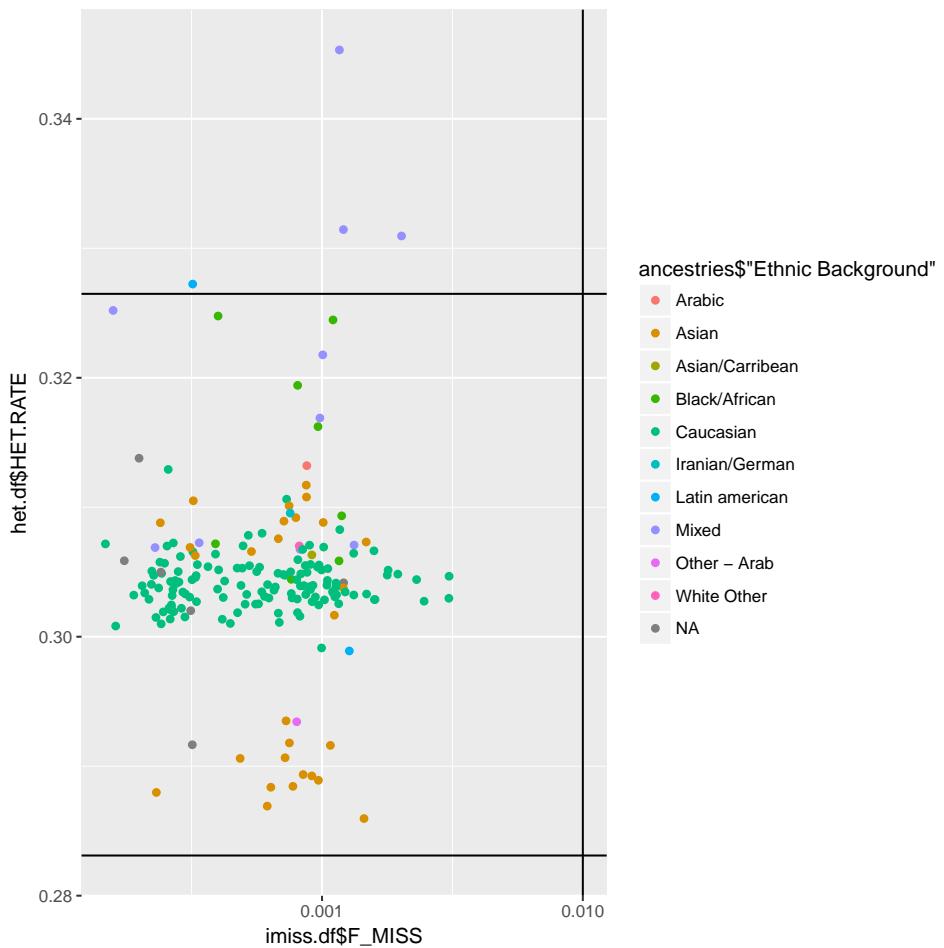


Figure 1.4: Sample filters for missingness and heterozygosity rate. Samples outside the central rectangle were excluded.

1.2.5 Computing genotype principal components as covariates for ancestry

As shown in [Table 1.1](#), the [HIRD](#) cohort is multi-ethnic, hence there is potential for confounding by population structure (sample structure due to genetic ancestry) in expression and genetic association studies^{26–28}. Treating HapMap 3 samples as a reference population where the major axes of variation in genotypes are likely to be ancestry, [principal component analysis \(PCA\)](#) was performed using smartpca (v8000) on linkage disequilibrium (LD)-pruned genotypes (PLINK --indep-pairwise 50 5 0.2). [HIRD](#) sample [principal components \(PCs\)](#) were computed by projection onto the HapMap 3 [PCA](#) eigenvectors. For non-genotyped individuals, [PC](#) values were imputed as the mean value for all genotyped individuals with the same self-reported ancestry. The top [PCs](#) separate samples of European, African and Asian ancestry ([Fig. 1.5](#)), hence these [PCs](#) can be used as covariates for ancestry downstream.

Add Tracy-Widom statistics for PCs to justify later choice of 4 PCs for covariates

1.2.6 Genotype phasing and imputation

Prior to imputation, 213277 monomorphic markers that provide no information for imputation were removed. Imputation for the autosomes and X chromosome was conducted using the Sanger Imputation Service*, which involves pre-phasing with EAGLE2 (v2.4), then imputation with PBWT (v3.1) using the Haplotype Reference Consortium (r1.1) panel. Markers were lifted-over from GRCh37 to GRCh38 coordinates using CrossMap. Poorly-imputed markers with INFO < 0.4 or missingness > 5% were removed, resulting in 40290981 markers.

1.2.7 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 batch effects from pooling, samples were pooled by library prep plate, ensuring libraries from all timepoints of an individual

*<https://imputation.sanger.ac.uk/>

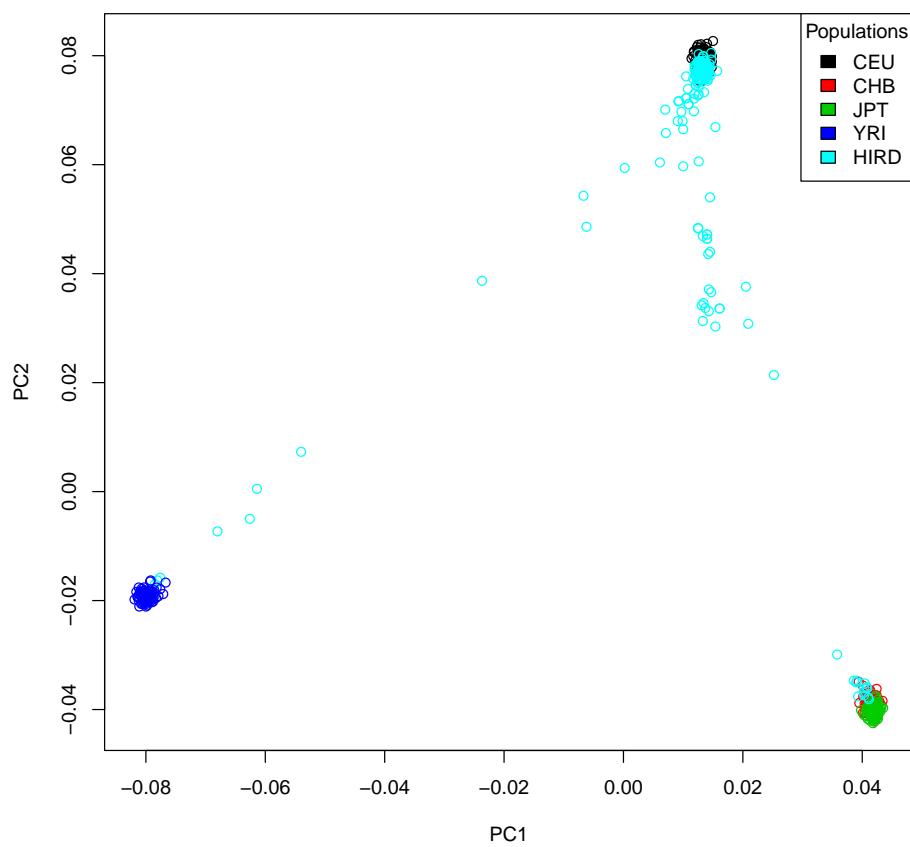


Figure 1.5: HIRD samples (cyan) projected onto PC1 and PC2 axes defined by PCA of HapMap 3 samples. The first two PCs separate European (CEU, upper-right) from Asian (CHB and JPT, lower-right) and African (YRI, lower-left) individuals.

were in the same pool, and then sequenced across multiple lanes as technical replicates (HiSeq 4000, 75bp paired-end).

Can add other fastqc plots e.g. kmers, overrepresented seqs, seq length

RNA-seq quality metrics were assessed using FASTQC* and Qualimap²⁹, then visualised with MultiQC³⁰. Sequence quality was high (Fig. 1.6), and duplication levels were low (Fig. 1.7). The unimodal GC-content distribution suggested negligible levels of non-human contamination (Fig. 1.8).

add software versions

1.2.8 RNA-seq quantification and filtering

Reads were quantified against the Ensembl reference transcriptome (GRCh38) using Salmon³¹ in quasi-mapping-based mode, which internally accounts for transcript length and GC composition. 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%), comfortably within sequencing depth recommendations for DGE experiments³². Relative transcript abundances were summarised to Ensembl gene-level count estimates using tximport (scaledTPM method) to improve statistical robustness and interpretability³³.

Genes with short noncoding RNA biotypes[†] were removed, as they are generally not polyadenylated, and expression estimates can be biased by

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

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

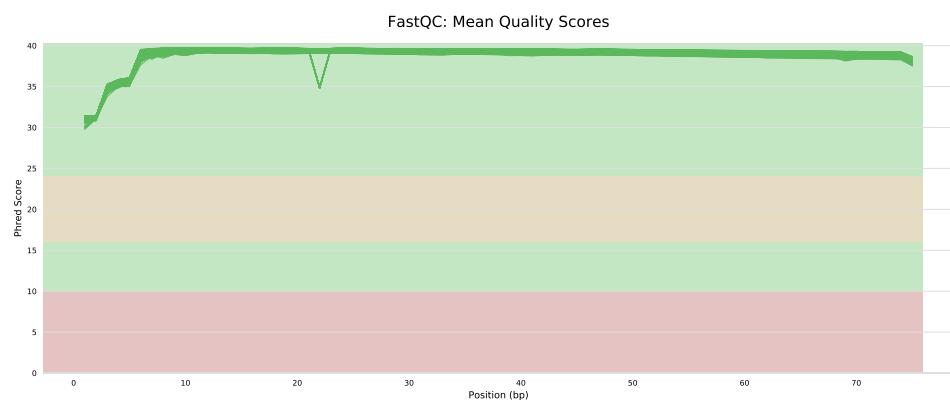


Figure 1.6: FastQC sequence quality versus read position for HIRD RNA-seq samples.

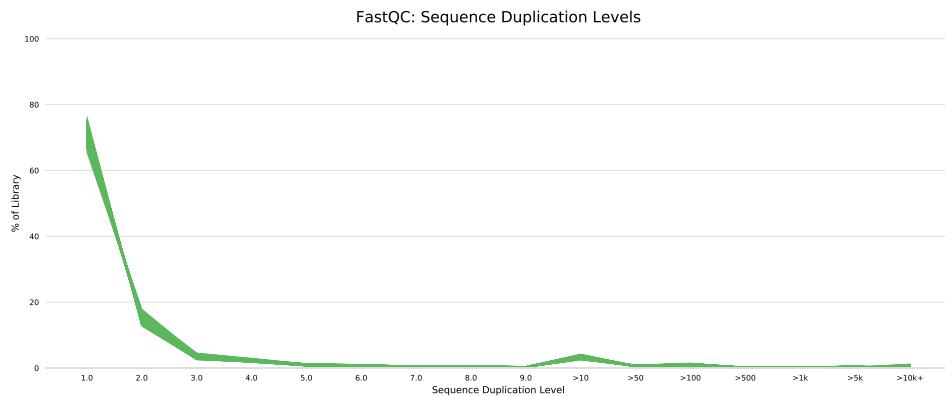


Figure 1.7: FastQC sequence duplication levels for HIRD RNA-seq samples.

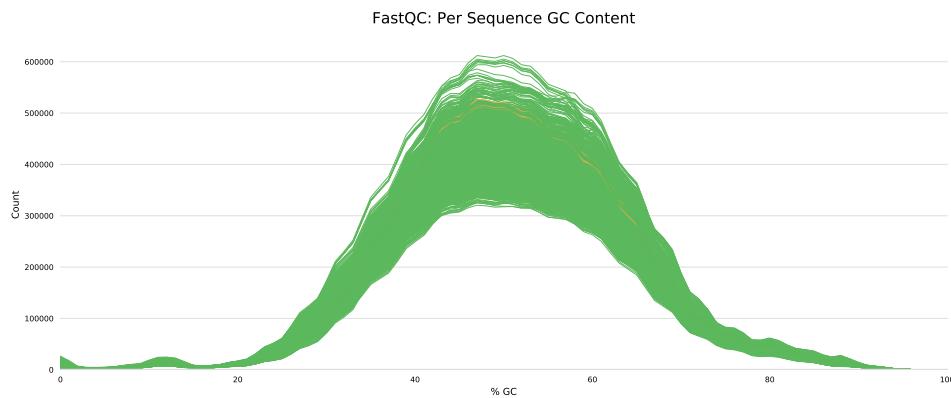


Figure 1.8: FastQC GC profile for HIRD RNA-seq samples.

misassignment of counts from overlapping protein-coding or lncRNA genes³⁴. Globin genes, which are highly expressed in erythrocytes and reticulocytes, cell types expected to be depleted in PBMC³⁵, were also removed. Given the proportion of removed counts at this stage was low for most samples (Fig. 1.9), poly(A) selection and PBMC isolation procedures were deemed to have been efficient.

Many of the genes in the reference transcriptome are not expressed in PBMC (Fig. 1.10), and many genes are expressed at counts too low for statistical analysis of DGE. Genes were further filtered to require detection (non-zero expression) in at least 95% of samples, and 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 expressed³⁶. The change in the distribution of gene expressions among samples before and after filtering shows a substantial number of low expression genes are removed (Fig. 1.11).

After the application of all filters, expression values were available for 21626 genes over 223 samples (75/75 individuals on day 0, 73/75 on day 1, and 75/75 on day 7).

1.2.9 Array data preprocessing

Single-channel Agilent 4x44K microarray (G4112F) data for 173 samples from¹⁸ were downloaded from ArrayExpress*. These arrays were originally processed in two batches, the effect of which is seen in the raw foreground intensities (Fig. 1.12).

VSN³⁷ was used to perform background correction, between-array normalisation, and variance-stabilisation of intensity values, resulting in expression values on a log₂ scale.

Most genes are targeted by multiple array probes; 31208 probes were collapsed into 18216 Ensembl genes using by selecting the probe with the highest mean intensity for each gene (`WGCNA::collapseRows(method=MaxMean)`, recommended for probe to gene collapsing³⁸). 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

*<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2313/>

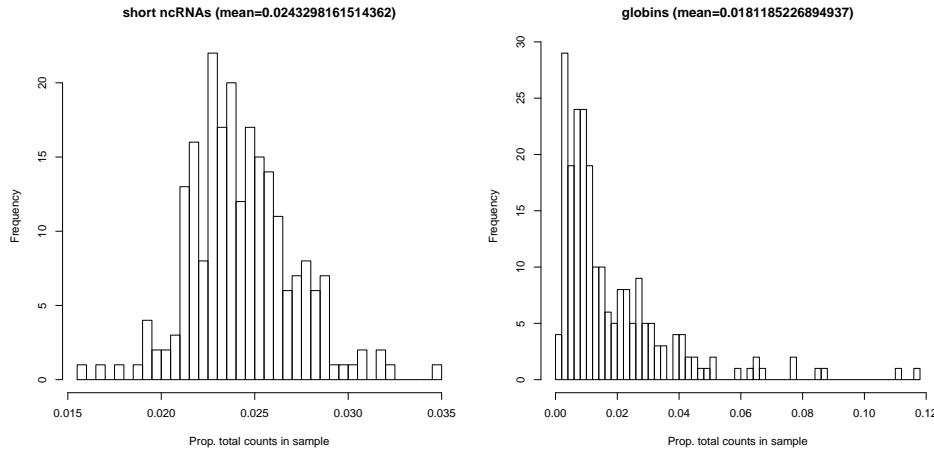


Figure 1.9: Distributions of removed short ncRNA and globin counts as a proportion of total counts in RNA-seq samples.

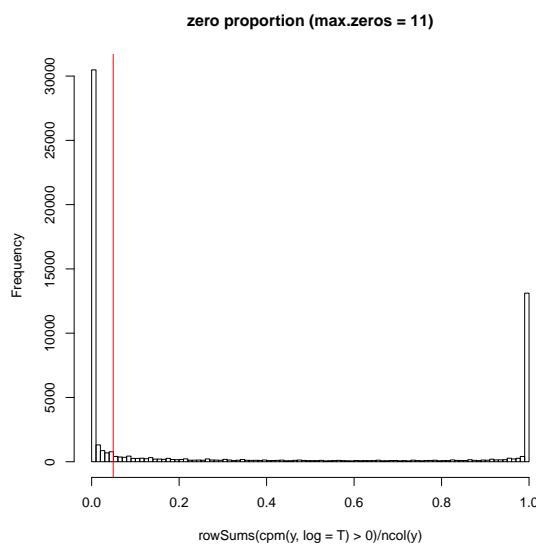


Figure 1.10: Distribution of the proportion of samples in which genes were detected (non-zero expression). Many genes are not detected in any samples. Vertical line shows 5% threshold below which genes were discarded.

CHAPTER 1. TRANSCRIPTOMIC RESPONSE TO INFLUENZA A
1.2. METHODS (H1N1)PDM09 VACCINE

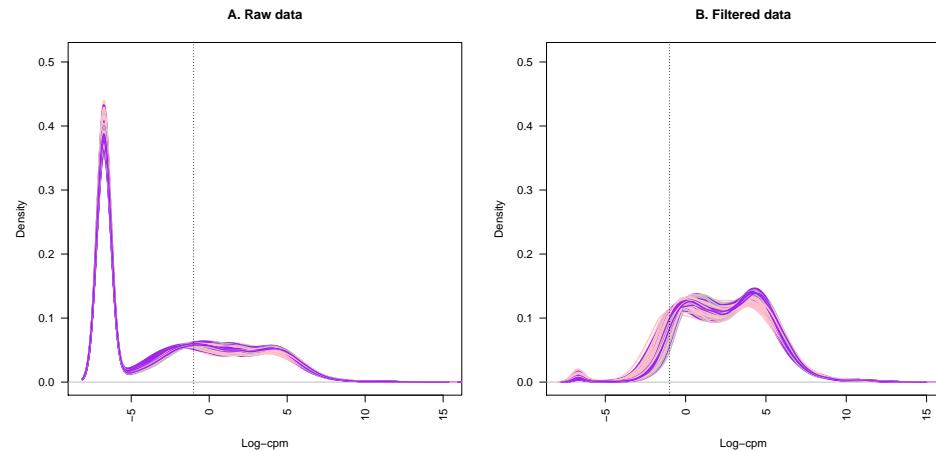


Figure 1.11: Distribution of gene expressions for RNA-seq samples before and after filtering no expression and low expression genes. Vertical line shown at CPM = 0.5 threshold.

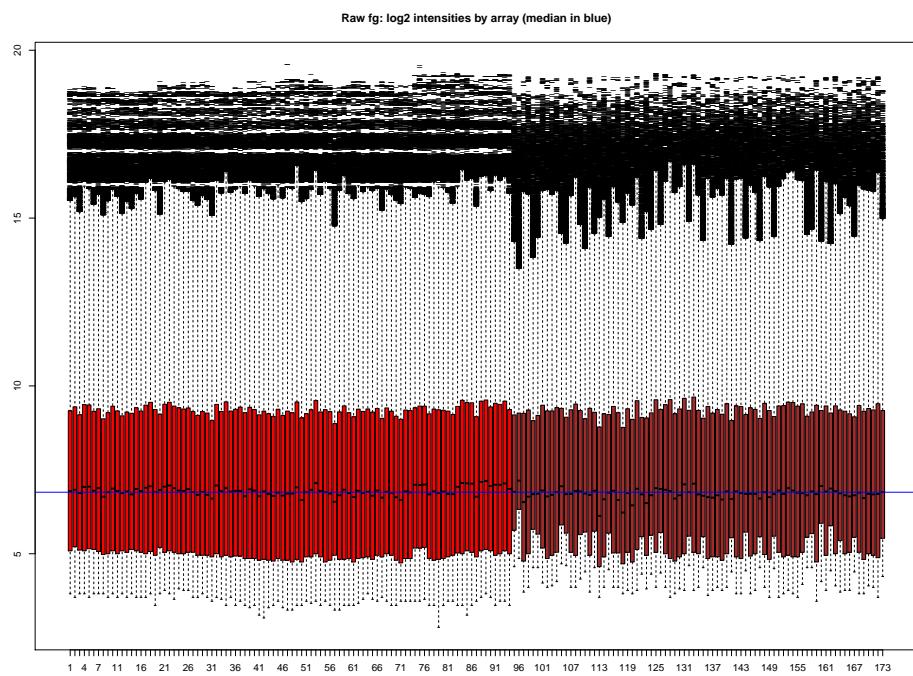


Figure 1.12: Raw foreground intensities for 173 HIRD array samples. Colored by array processing batch.

in the **HIRD** dataset. The final normalised \log_2 intensity values for these 18216 genes over 173 samples is shown in [Fig. 1.13](#).

1.2.10 Differential gene expression

PCA of the expression data reveals although samples separate by experimental timepoint along PC3 ([Fig. 1.14d](#)), measurement platform is by far the largest source of variation. Normalisation was also not able to completely remove the batch effect within the array data ([Fig. 1.14a](#)). The large platform effect likely stems from systematic technological differences in how each platform measures expression. For example, arrays suffer from ratio compression due to cross-hybridisation³⁹. **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 intensity⁴⁰. There are also differences in the statistical models behind expression quantification and normalisation, as described above.

Despite the shortcomings of array data detailed above, the array dataset tends to contain individuals with more extreme antibody response phenotypes ([Fig. 1.3](#)), and hence the data should not be excluded. Given the magnitude of the platform effect, I concluded that the appropriate approach should be a two-stage approach that integrates 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 ComBat⁴¹, which estimates centering and scaling parameters by pooling information across all genes using empirical Bayes. ComBat is the method used in¹⁸. In comparisons of microarray batch effect adjustment methods, ComBat performs favourably (vs. five other adjustment packages)⁴² or comparably (vs. batch as a fixed or random effect in the linear model)⁴³. However, where batches are unbalanced in terms of sample size⁴⁴ or distribution of study groups that have an impact on expression⁴⁵, 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, but the proportion of responders is not [Table 1.2](#), hence I elect not to use ComBat to pre-adjust the array expression data, and model the batches as fixed effects. In practice, results from the **DGE** analysis were not substantially affected by the choice of whether to use a ComBat pre-adjustment or a fixed effect.

cite relevant preprocessing sections

combat does have a pro in that it can do per gene scaling, that fixed fx won't do

this is not a very precise justification. actually, if I were to color R/NR in the PCA plot, R/NR doesn't really explain a lot of var in global gene expression. that's probably why the results don't change much.

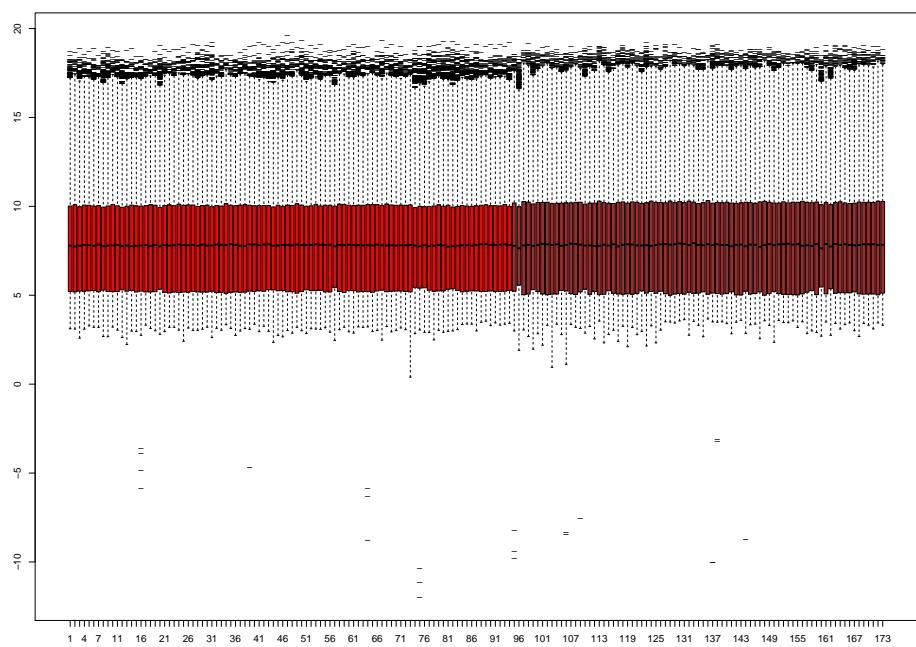


Figure 1.13: Array intensity estimates after VSN normalisation and collapsing of probes to genes. Colored by array processing batch.

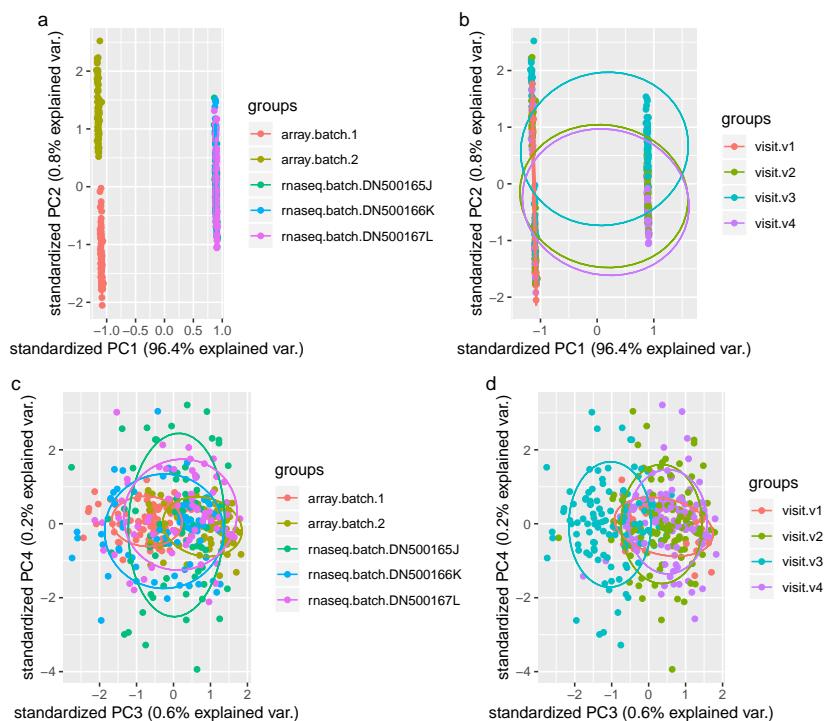


Figure 1.14: First four PCs in the HIRD expression data, colored by platform and batch (left), and timepoint (right).

1.2.10.1 Per-platform differential gene expression model

For the array data, as¹⁸ demonstrated no significant global differences in expression between day -7 and day 0, I likewise merge these two timepoints into a single “day 0” baseline timepoint in the following DGE models.

For the RNA-seq data, between-sample normalisation was performed using the trimmed mean of M-values (TMM) method⁴⁶ from edgeR⁴⁷; then variance-stabilisation was performed using voom⁴⁸, resulting in expression values with units of log₂ CPM.

this is DGE specific normalisation, which is why it goes here, not in the preprocessing section

Linear models were fit using limma⁴⁹, which is computationally fast, and performs well for sufficiently large ($n \geq 3$ per group) sample sizes⁵⁰. For each gene, I fit a model (model 1) with expression as the response variable; with timepoint (baseline, day 1, day 7), TRI, batch, 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, including 3 additional terms for the interactions between each timepoint and TRI. From model 1, I defined contrasts for day 1 vs. baseline, day 7 vs. baseline, day 7 vs. day 1, TRI, sex, and age. From model 2, I defined contrasts for the TRI specifically at each of the three timepoints. Corresponding coefficients and standard errors for the contrasts were extracted from the linear models, which represent effect size in units of log₂ expression fold change per unit change in predictor value.

link to papers justifying sex, age, ancestry as significant effects on immune gene expression

1.2.10.2 Choice of differential gene expression meta-analysis method

add section labels

In the section , I concluded that a two-stage meta-analysis approach would be appropriate. This meta-analysis is restricted to 13593 genes assayed by both the array and RNA-seq platforms.

Two popular frameworks for effect size meta-analysis are fixed-effect and random-effects^{51,52}. Given k studies, the fixed-effect model assumes a common population effect size shared across all studies, with observed variation explained only by sampling error. The random-effects model assumes the k study-specific effect sizes are drawn from some distribution with variance τ^2 (standard deviation (SD) τ), representing an additional source of variation termed the between-studies heterogeneity, reducing to the fixed-effect model when $\tau = 0$. In the HIRD data, there are $k = 2$ ‘studies’ (array and RNA-seq),

where the platform differences described in section contribute to considerable between-studies heterogeneity. The assumption of $\tau = 0$ is unrealistic, hence 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 ⁵⁴, in the case of $k = 2$ especially⁵⁵. Many estimators are available⁵⁶, 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^{57,58}. In such circumstances, the most sensible choice may be to incorporate prior information about model hyperparameters in a Bayesian random-effects framework^{56–59}. For this study, I use the implementation in bayesmeta⁵³, which requires priors for both effect size and between-studies heterogeneity.

add label

make all the notation in this section consistent with, and add the equation 2.1. The normal-normal hierarchical model,⁵³

1.2.10.3 Prior for between-studies heterogeneity

The choice of prior for between-studies heterogeneity is influential when k is small⁵⁹. Gelman [60] 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. Since I assumed zero estimates for τ are unrealistic, I use a weakly-informative gamma prior recommended by⁵⁷, which has zero density at $\tau = 0$, increasing gently as τ increases. 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.^{61,62}) or inverse-gamma (e.g.⁶³) families that have derivatives or zero close to zero, thus ruling out small values of τ no matter what the data suggest; and in contrast to half-t family priors (e.g.^{59,60}), which have their mode at zero, and do not rule out $\tau = 0$.

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 effects⁵⁶) was first for each gene using `metafor::rma`. Genes with small estimates of $\tau < 0.01$ were excluded, and a gamma distribution was fit to the remaining estimates using `fitdistrplus`.

1.2.10.4 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^{58,60}. `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 gene-wise 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 tails⁶⁴. Since `bayesmeta` does not implement a Cauchy prior, to avoid over-shrinkage, I flatten the normal prior considerably by scaling up the variance to $N(0, 100\sigma^2)$. This is equivalent to assuming placing a 95% prior probability that effects are less extreme than approximately 20σ .

1.2.10.5 Evaluation of priors

An example of the empirically estimated hyperparameters for the priors for the day 1 vs. baseline contrast are shown in Fig. 1.15 (for τ) and Fig. 1.16 (for μ). For τ , the final prior used was $\text{Gamma}(\text{shape} = 1.5693, \text{scale} = 0.0641)$. This is comparable to⁵⁷'s default recommendation of a $\text{Gamma}(\text{shape} = 2, \text{scale} = \lambda)$ prior where λ is small. For μ , the final prior used was $N(0, (0.3240 * 10)^2)$. 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.

1.2.10.6 Multiple testing correction

For the frequentist random-effects meta-analysis, nominal gene-wise p values are converted to **false discovery rate (FDR)** estimates using the **Benjamini-Hochberg (BH)** procedure (`p.adjust` in R). For the Bayesian random-effects meta-analysis, posterior effect sizes and standard errors are supplied to `ashr`, which estimates the **local false sign rates (lfsrs)**, which are analogous to **FDR**, but quantifies the probability of calling the wrong sign for an effect rather than the confidence of a non-zero effect⁶⁵.

why is this? is it having well powered studies? gelman is vague

the derivation here is
 $\text{qnorm}(0.975, \text{mean}=0, \text{sd}=1*10) = 1*19.59964$,
 bit iffy, double check this is correct

could also include a table of all sets of parameters here?

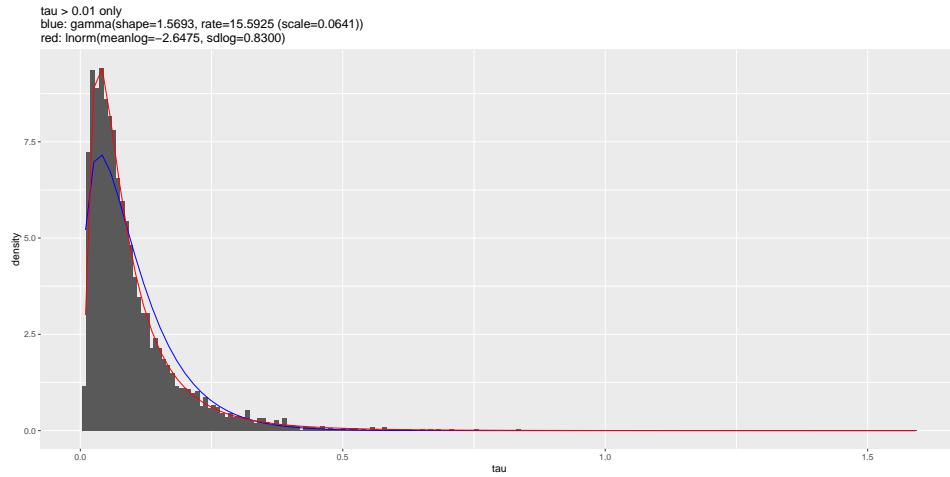


Figure 1.15: Gamma prior for τ used for `bayesmeta` (blue), compared to the empirical distribution of per-gene frequentist `metafor::rma` estimates for τ , for the day 1 vs. baseline effect (small estimates of $\tau < 0.01$ excluded). Empirical log-normal fit also shown (red).

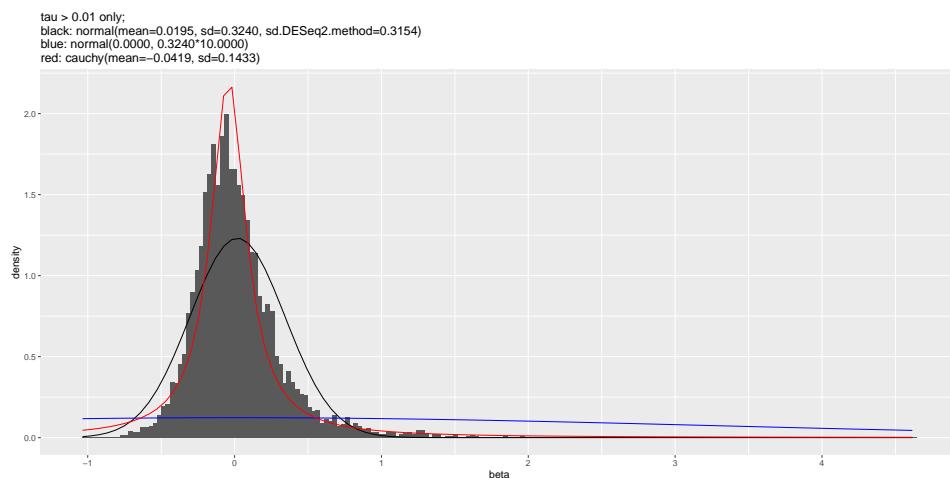


Figure 1.16: Normal prior for μ used for `bayesmeta` (blue), compared to the empirical distribution of per-gene frequentist `metafor::rma` estimates for τ , for the day 1 vs. baseline effect. The non-scaled normal fit is shown (black), as well as a Cauchy fit (red).

1.2.11 Gene set enrichment analysis using blood transcription modules

Gene set enrichment analyses were conducted using `tmod::tmodCERNOtest`⁶⁶, which assesses the enrichment of small ranks within specific sets of genes compared to all genes, when the genes are ranked by some metric—here I used effect sizes from `bayesmeta`. The gene sets used were **blood transcription modules (BTMs)** from⁶⁷, which are annotated sets of coexpressed genes mined from publicly available human blood transcriptomic data, and provide sets tailored for enrichment analyses in blood cells.

1.3 Results

1.3.1 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 13593 genes with expression measured by both platforms, models were fit within each platform, then effect sizes were combined using Bayesian random-effects meta-analysis.

At a **lfsr** < 0.05 and absolute **FC** > 1.5 cutoff, 857/13593 genes were differentially expressed between any pair of timepoints, with their expression clustering into three main clusters (Fig. 1.17).

1.3.2 Innate immune response at day 1 post-vaccination

Consistent with global expression at day 1 being markedly different from expression at other timepoints (Fig. 1.14), the highest numbers of differentially expressed genes are 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* ($\log_2 \text{FC} = 4.489\,150$), an interferon-induced gene in monocytes and **dendritic cells (DCs)** involved in antiviral innate immune pathways⁶⁸. Other key genes in the interferon signalling pathway⁶⁹ such as *STAT1* ($\log_2 \text{FC} = 2.169\,3060$), *STAT2* ($\log_2 \text{FC} = 0.948\,9341$), and *IRF9* ($\log_2 \text{FC} = 0.815\,3674$) are also

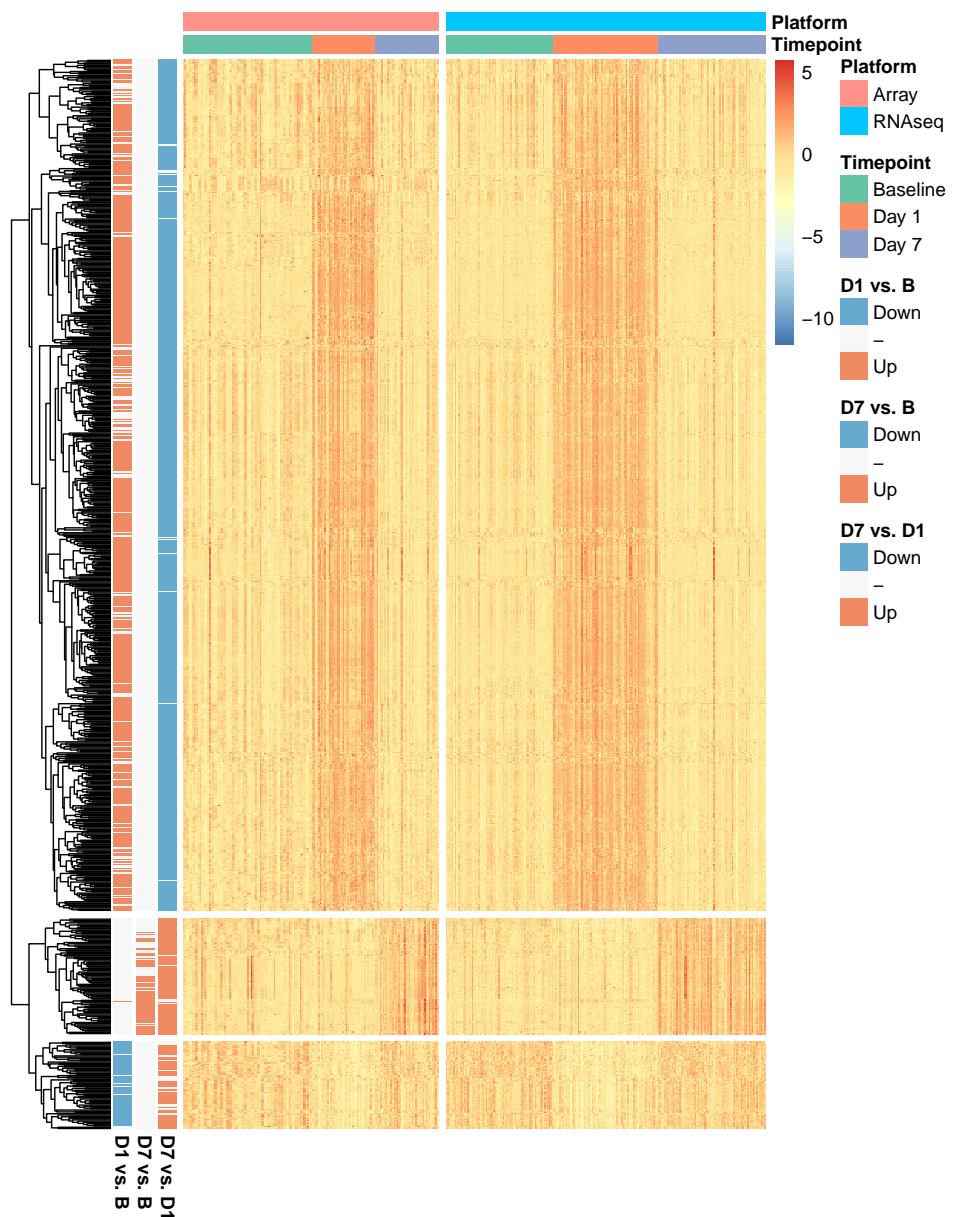


Figure 1.17: Normalised gene expression for genes differentially expressed between any pair of timepoints ($\text{lfsr} < 0.05$, absolute fold change > 1.5) across HIRD samples, clustered by gene (Manhattan distance metric).

upregulated at day 1. Gene set enrichment analysis using `tmod` revealed that genes with the high **FC** increases at day 1 were enriched in modules associated with activated **DCs**, monocytes, toll-like receptor and inflammatory signalling (Fig. 1.18), confirming that day 1 responses are dominated by signatures of innate immunity. 64 genes were downregulated at day 1, enriched in modules associated with T cells and **natural killer (NK)** cells, with the largest absolute fold change observed for *FGFBP2* ($\log_2 \text{FC} = -0.9141547$). For both up and downregulated genes, there was a tendency to return to baseline expression levels by day 7.

1.3.3 Adaptive immune response at day 7 post-vaccination

59 genes were differentially expressed at day 7 vs. baseline, with expression fold changes more modest than those at day 1. The genes with the highest up-regulation were the B cell-associated genes *TNFRSF17* ($\log_2 \text{FC} = 1.7538617$) and *MZB1* ($\log_2 \text{FC} = 1.7369668$). Plasma cell-specific genes including *SDC1* (encodes CD138 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5437827/>) ($\log_2 \text{FC} = 1.3673081$) and *ELL2* (<https://www.nature.com/articles/ni.1786>) ($\log_2 \text{FC} = 0.8679659$) were also prominently upregulated. Strongly enriched modules at day 7 were related to mitosis and cell proliferation, particularly in CD4^+ T cells (Fig. 1.18). Both the CD4^+ T cell and plasma cell response are indications of an adaptive immune response at day 7.

1.3.4 Expression signatures associated with antibody response

I also looked for genes which have expression associated with baseline-adjusted antibody response, as quantified by **TRI**. At the initial frequentist meta-analysis stage, with a significance threshold of **FDR** < 0.05, 6 genes had expression associated with **TRI** at baseline, 55 at day 7, and 11 pooling samples across timepoints (Fig. 1.19).¹⁸ also identified genes with day 7 expression associated with antibody response, where response was defined as a binary phenotype based on 4-fold change (described in section). They reported 62 significant associations at **FDR** < 0.05, of which 58/62 fall into the 13593 genes considered in my meta-analysis (circled, Fig. 1.19), and 15/58 replicated, all with the same positive direction of effect (high expression with high **TRI**). In the Bayesian meta-analysis, no single gene was detected as

can also add MSigDB hallmark sets, which include interferon sets; and of course gene ontology sets

not sure of interpretation at FGFBP2, it is indeed highly expressed in NKs through <https://dice-database.org/genes/FGFBP2>

any point in a table of e.g. top 20 DE genes, or is the gene set analysis already enough?

change x axis labels to baseline, specify top 10 procedure in figure caption

finish citing

add label

CHAPTER 1. TRANSCRIPTOMIC RESPONSE TO INFLUENZA A (H1N1)PDM09 VACCINE

1.3. RESULTS

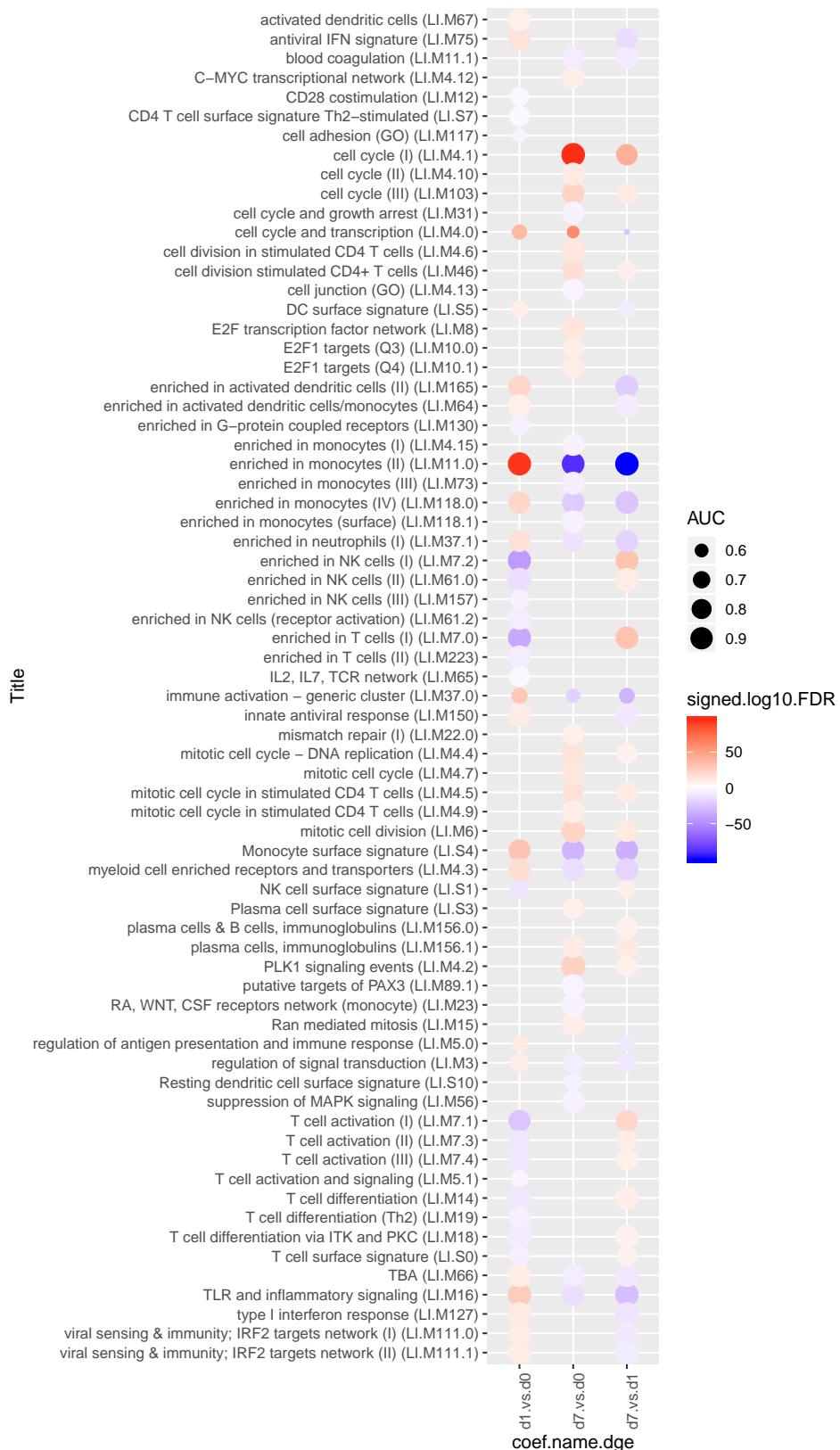


Figure 1.18: Transcriptomic modules significantly up or downregulated post-vaccination. Size of circle indicates effect size. Color of circle indicates significance and direction of effect (red = upregulation, blue = downregulation).

significantly associated with **TRI** at $\text{lfsr} < 0.05$ at any timepoint, or when pooling samples across all timepoints (Fig. 1.20).

Significant enrichments were detected at the gene set level; the strongest effects are seen at day 7, where expression of cell cycle, CD4⁺ T cells, and plasma cells are associated with high **TRI**. At day 0, modules related with inflammatory response in myeloid cells are also associated with high **TRI** (Fig. 1.21).

figure x labels here should
be TRI, not R.vs.NR

1.3.5 Identifying expression signatures for predicting antibody response [probably cut this section and just add to discussion]

1.4 Discussion

There is extensive transcriptomic response to Pandemrix vaccination in the **HIRD** cohort. Upregulation of genes and modules related to the interferon signalling pathway, monocytes, inflammatory response, and other aspects of innate immunity were detected at day 1. This response is transient, with most such genes returning to baseline expression by day 7. Upregulation of cell cycle/proliferation, activated CD4⁺ T cell, and B (plasma) cell genes and modules were detected at day 7. This is likely a signature indicating the shift to an adaptive immune response, involving CD4⁺ T cell-supported differentiation and proliferation of antibody-secreting plasmablasts and plasma cells⁷¹. These patterns of expression change between timepoints in the **RNA-seq** data are consistent with the patterns in the array data in the original study¹⁸, and with expansions of monocyte and plasma cell populations seen in the **FACS** data at days 1 and 7 respectively in the original **HIRD** study¹⁸.

In contrast, I was not able to fully replicate the originally reported single gene-level associations between day 7 expression and antibody response in the **RNA-seq** data and subsequent and meta-analyses. In¹⁸, 62 genes were reported as differentially expressed between vaccine responders and non-responders. Although¹⁸ encodes responder status as a binary phenotype, whereas my analysis uses **TRI**, this is not the primary difference, as 51/62 genes replicated ($\text{FDR} < 0.05$) using **TRI** when considering just the array data. The same analysis using only the **RNA-seq** data replicated 0/62 genes.

The majority of the effects for these genes were simply much stronger in the array dataset than in the RNAseq dataset (Fig. 1.19). Given that the

Not sure if there is a biological interpretation of downreg of T cells and NK cells gene sets at day 1, since it could be due to increase in other cell types in the sample. similar findings in⁷⁰ though

might have to rerun everything using the original binary R/NR if this line of reasoning isn't strong enough

move numbers to results?

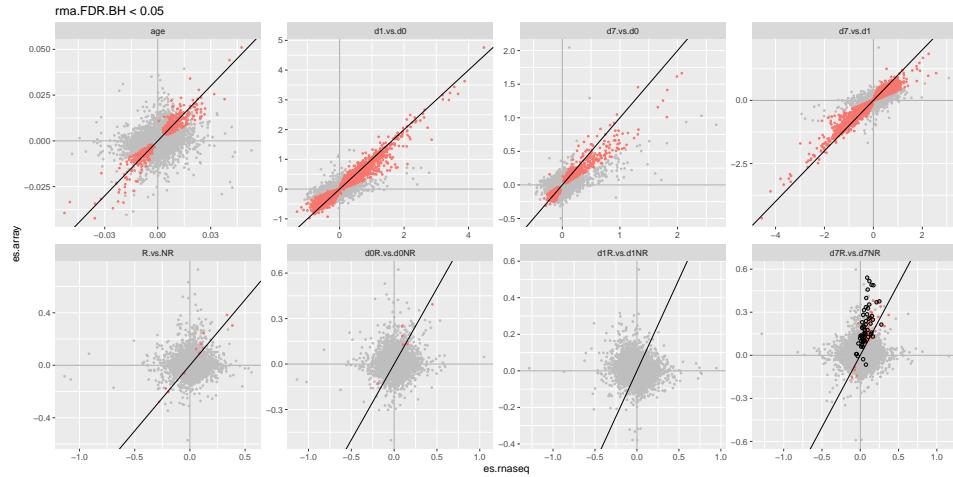


Figure 1.19: DGE effect sizes estimated in array vs. RNA-seq. Significance colored by frequentist random effects meta-analysis FDR < 0.05. Genes with day 7 expression associated with responder/non-responder status in¹⁸ are circled for that contrast.

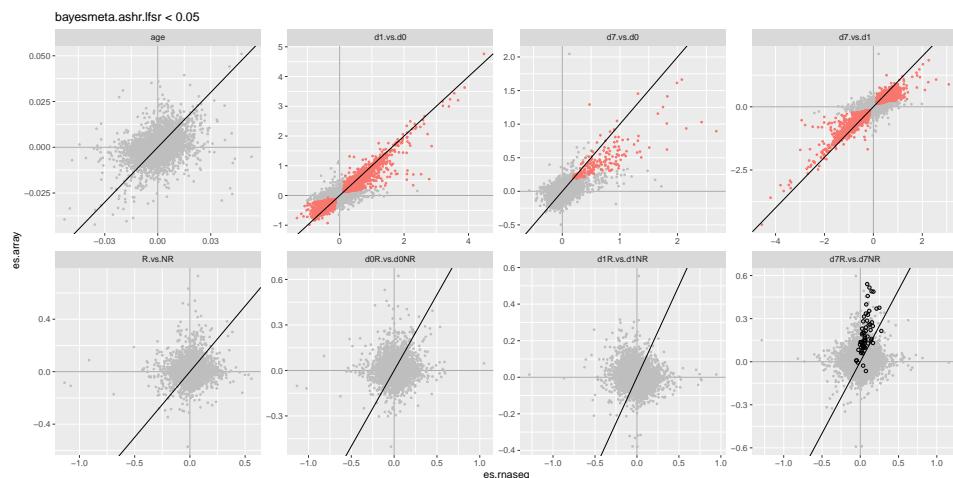


Figure 1.20: DGE effect sizes estimated in array vs RNA-seq. Significance colored by Bayesian random effects meta-analysis lfsr < 0.05. Genes with day 7 expression associated with responder/non-responder status in¹⁸ are circled for that contrast.

CHAPTER 1. TRANSCRIPTOMIC RESPONSE TO INFLUENZA A
1.4. DISCUSSION (H1N1)PDM09 VACCINE



Figure 1.21: Transcriptomic modules enriched in genes with expression associated with antibody response (**TRI**) at each day. Size of circle indicates effect size. Color of circle indicates significance and direction of effect (red = expression positively correlated with TRI, blue = negative).

range of **TRI** is higher in the array individuals (Table 1.1), this does not seem unusual that stronger **TRI**-associated effects are observed there.

58/62 reported hits were measured by both platforms and assessed in the meta-analysis. Only 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 **REML** estimate of between-platform heterogeneity was zero for 8563/13593 for the day 7 **TRI** contrast overall, and zero for all 15 of these signals. None of these signals replicated in the Bayesian random-effects meta-analysis. The Bayesian meta-analysis is in general more conservative, calling fewer differentially expressed genes compared to the frequentist analysis for all contrasts (Fig. 1.20). Prior information about τ is incorporated, discouraging unrealistic estimates of zero heterogeneity. Given the between-platform heterogeneity coming from both platform-specific technical differences and **TRI** phenotype differences, relative to the modest effect size distributions compared to between-timepoint **DGE** comparisons, the data are not well-positioned to identify significant single-gene associations with antibody response.

Expression signatures of antibody response were, however, observed at the gene set level, for modules of coexpressed genes that are associated with **TRI** as a whole. The strongest effects were observed at day 7, where expression of adaptive immune response modules (cell cycle, stimulated CD4⁺ cell, plasma cell modules) were positively associated with **TRI**. These are the same modules observed to be upregulated at day 7 compared to baseline; it seems that those individuals with the greatest antibody response to vaccination are most able to upregulate these gene sets by day 7 post-vaccination.

Module associations were also observed pre-vaccination (cell adhesion, enriched in B cells, proinflammatory cytokines, platelet activation), suggesting baseline immune state has some influence on long-term antibody response to Pandemrix. Over the years, a diverse range of gene sets have been found to be baseline predictors of serological response to influenza vaccination: apoptosis¹⁹; Fc γ receptor-mediated phagocytosis, TREM1 signaling²⁰; enriched in B cells, T cell activation²¹; B cell receptor signalling, inflammatory response, platelet activation²²; several of which I also observe. It should be noted that comparisons with these signatures from existing influenza systems vaccinology studies should be caveated, as most existing studies are for non-adjuvanted influenza vaccines. Adjuvanted influenza vaccines are

CHAPTER 1. TRANSCRIPTOMIC RESPONSE TO INFLUENZA A
1.4. DISCUSSION (H1N1)PDM09 VACCINE

could comment on phenotype differences too, i.e. HIRD measure antibodies at d63, much later than is popular in the field: d28 usually

should probably emphasize sobolev didn't find pre-vacc signatures, and we did. But it's not exactly fair, as sobolev didn't use gene set enrichment as far as i can tell

There is also something to be said about 'prediction is not inference'. For use as correlates of protection, as promised by proponents of systems studies, prediction is what is important.

found signatures, but so what? Feels like chapter lacks a punchline?

considerably more immunogenic, and post-vaccination expression patterns differ to those of non-adjuvanted vaccines^{16,18}. Hence, it is particularly important that the robustness of these observed baseline expression signatures be validated in an independent cohort for a comparable AS03-adjuvanted influenza vaccine.

In conclusion, Chapter 2 characterises the expansive changes in **PBMC** gene expression that follow vaccination with Pandemrix. The dominant trend for all individuals is transient upregulation of the innate immune response at day 1, transitioning into adaptive immunity by day 7. Baseline-adjusted antibody response is correlated with expression of gene sets, particularly adaptive immunity modules at day 7, but also for some modules pre-vaccination. Unfortunately, between-platform variation in expression impedes identification of specific genes that contribute. The fundamental question of why gene expression and antibody responses vary between **HIRD** individuals remains. Chapter 3 will examine one hypothesis: the impact of common human genetic variation on Pandemrix expression response.

Table 1.1: Sample descriptive statistics.

	Total n = 114	array n = 44	platform rnaseq 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)
Ethnic Background			
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 HAI 0	4.4 (1.8)	4.2 (1.6)	4.5 (1.9)
log2 HAI 6	7.6 (1.8)	7.4 (2.2)	7.6 (1.5)
log2 HAI ratio	3.2 (1.9)	3.2 (2.4)	3.1 (1.6)
log2 MN 0	6.2 (2.8)	5.4 (2.4)	6.6 (3.0)
log2 MN 6	10.4 (2.0)	9.5 (2.2)	10.9 (1.6)
log2 MN ratio	4.2 (2.3)	4.1 (2.6)	4.3 (2.1)
responder			
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)

CHAPTER 1. TRANSCRIPTOMIC RESPONSE TO INFLUENZA A
1.4. DISCUSSION *(H1N1)PDM09 VACCINE*

Table 1.2: HIRD batch balance

	Total n = 374	1 n = 87	2 n = 79	batch DN500165J n = 70	DN500166K n = 69	DN500167L n = 69
visit						
v1	40 (10.7%)	20 (23%)	20 (25.3%)	0 (0%)	0 (0%)	0 (0%)
v2	114 (30.5%)	24 (27.6%)	20 (25.3%)	24 (34.3%)	23 (33.3%)	23 (33.3%)
v3	109 (29.1%)	21 (24.1%)	20 (25.3%)	22 (31.4%)	23 (33.3%)	23 (33.3%)
v4	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)

Appendix A

Supplementary Materials

A.1 Chapter 2

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

A.2 Chapter 3

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus

luctus mauris.

A.3 Chapter 4

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Bibliography

1. 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) (Dec. 2018).
2. 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) (Mar. 2015).
3. Sautto, G. A., Kirchenbaum, G. A. & Ross, T. M. Towards a Universal Influenza Vaccine: Different Approaches for One Goal. *Virology Journal* **15**. doi:[10.1186/s12985-017-0918-y](https://doi.org/10.1186/s12985-017-0918-y) (Dec. 2018).
4. 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) (Oct. 2011).
5. Klimov, A. *et al.* in *Influenza Virus* (eds Kawaoka, Y. & Neumann, G.) Series Title: Methods in Molecular Biology, 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).
6. Plotkin, S. A. Correlates of Protection Induced by Vaccination. *Clinical and Vaccine Immunology* **17**, 1055–1065. doi:[10.1128/CVI.00131-10](https://doi.org/10.1128/CVI.00131-10) (July 2010).
7. 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) (Feb. 2013).
8. 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).

APPENDIX A. BIBLIOGRAPHY

9. 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).
10. 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) (Jan. 2019).
11. 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) (Apr. 2010).
12. 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) (Apr. 2011).
13. 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) (July 10, 2011).
14. 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) (Jan. 2014).
15. 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) (Mar. 2012).
16. 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) (Dec. 13, 2017).
17. 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) (Mar. 4, 2018).

APPENDIX A. BIBLIOGRAPHY

18. 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) (Jan. 4, 2016).
19. 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) (Apr. 16, 2013).
20. 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) (Apr. 10, 2014).
21. 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) (Dec. 2015).
22. 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) (Aug. 25, 2017).
23. Cohen, J. The Cost of Dichotomization. *Applied Psychological Measurement* **7**, 249–253. doi:[10.1177/014662168300700301](https://doi.org/10.1177/014662168300700301) (June 1983).
24. 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) (Jan. 2009).
25. Food and Drug Administration. *Guidance for Industry: Clinical Data Needed to Support the Licensure of Pandemic Influenza Vaccines* (Jan. 6, 2007), 20.
26. 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) (Aug. 2006).
27. 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) (July 17, 2014).

APPENDIX A. BIBLIOGRAPHY

28. Brown, B. C., Bray, N. L. & Pachter, L. Expression Reflects Population Structure. doi:[10.1101/364448](https://doi.org/10.1101/364448) (July 8, 2018).
29. 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) (Oct. 1, 2015).
30. 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) (Oct. 1, 2016).
31. 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) (Apr. 6, 2017).
32. 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) (Feb. 1, 2014).
33. 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) (Feb. 29, 2016).
34. 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) (Dec. 2018).
35. 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).
36. 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) (Aug. 2, 2016).

APPENDIX A. BIBLIOGRAPHY

37. 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 July 1, 2002).
38. 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).
39. 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) (Feb. 2006).
40. 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) (June 30, 2015).
41. 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) (Jan. 1, 2007).
42. 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) (Feb. 28, 2011).
43. 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) (Aug. 30, 2018).
44. 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) (Dec. 2018).
45. Nygaard, V., Rødland, E. A. & Hovig, E. Methods That Remove Batch Effects While Retaining Group Differences May Lead to Exaggerated

APPENDIX A. BIBLIOGRAPHY

- Confidence in Downstream Analyses. *Biostatistics*, kxv027. doi:[10.1093/biostatistics/kxv027](https://doi.org/10.1093/biostatistics/kxv027) (January Aug. 13, 2015).
46. 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) (Sept. 28, 2018).
 47. 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) (Jan. 1, 2010).
 48. 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).
 49. 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) (Apr. 20, 2015).
 50. 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) (Dec. 2013).
 51. 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).
 52. 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) (Apr. 2010).
 53. Röver, C. Bayesian Random-Effects Meta-Analysis Using the Bayesmeta R Package (Nov. 23, 2017).
 54. 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) (Apr. 6, 2018).

APPENDIX A. BIBLIOGRAPHY

55. 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) (July 20, 2015).
56. 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) (Mar. 2016).
57. Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A. & Liu, J. A Non-degenerate 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) (Oct. 2013).
58. 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) (Mar. 2017).
59. 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) (Dec. 2019).
60. 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) (Sept. 2006).
61. 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) (Nov. 20, 2011).
62. 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: R. M. TURNER ET AL. *Statistics in Medicine* **34**, 984–998. doi:[10.1002/sim.6381](https://doi.org/10.1002/sim.6381) (Mar. 15, 2015).
63. 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) (Dec. 30, 1996).

APPENDIX A. BIBLIOGRAPHY

64. 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) (June 1, 2019).
65. Stephens, M. False Discovery Rates: A New Deal. *Biostatistics*, kxw041. doi:[10.1093/biostatistics/kxw041](https://doi.org/10.1093/biostatistics/kxw041) (Oct. 17, 2016).
66. 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).
67. 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) (Dec. 15, 2013).
68. 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**, 201.4 LP –201.4 (1 Supplement May 1, 2016).
69. 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) (Mar. 21, 2014).
70. 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) (Feb. 16, 2016).
71. Murphy, K. & Weaver, C. *Janeway's Immunobiology* 9th edition. 904 pp. (Garland Science/Taylor & Francis Group, LLC, New York, NY, 2016).

List of Abbreviations

BH Benjamini-Hochberg

BTM blood transcription module

CPM counts per million

DC dendritic cell

DGE differential gene expression

FACS fluorescence-activated cell sorting

FC fold change

FDR false discovery rate

HA haemagglutinin

HAI haemagglutination inhibition

HIRD Human Immune Response Dynamics

LAI live attenuated influenza vaccine

LD linkage disequilibrium

lfsr local false sign rate

MAF minor allele frequency

MN microneutralisation

NA neuraminidase

NK natural killer

PBMC peripheral blood mononuclear cell

PC principal component

PCA principal component analysis

REML restricted maximum likelihood

RNA-seq RNA-sequencing

SD standard deviation

TMM trimmed mean of M-values

TRI titre response index